.

Database Development for Model Inputs to Assess California Methyl Bromide Ambient Concentrations

January 20, 2011

Prepared for

Loren Hall
U.S. Environmental Protection Agency
Office of Civil Rights (1201A)
1200 Pennsylvania Ave., NW
Washington, DC 20460

Prepared by

ICF Consulting, Inc.
9300 Lee Highway
Fairfax, VA 22031

Contract Number EP-C-09-009

DISCLAIMER

Reference in this documentation to any of the following does not necessarily constitute or imply the endorsement, recommendation, or favoring by EPA or the United States Government or their employees or contractors: (1) specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise or (2) other documents, external web sites, or other references and the information contained therein. Furthermore, the views and opinions of authors expressed herein do not necessarily state or reflect those of EPA or the United States Government or their employees or contractors, and shall not be used for advertising or product endorsement purposes. Neither EPA nor the United States Government nor their employees or contractors makes any warranty, expressed or implied, including the warranties of merchantability and fitness for a particular purpose, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights, with respect to documents referenced in this document.

Table of Contents

List of Figures

[This page intentionally left blank]

1 INTRODUCTION

The EPA Office of Civil Rights (OCR) is responsible for ensuring compliance with the provisions of Title VI of the Civil Rights Act and EPA's implementing regulations. These provisions require that recipients of federal financial assistance operate their programs in a non-discriminatory manner. Persons may file administrative complaints with EPA identifying alleged actions by recipients that result in an adverse disparate impact on the basis of race, color, or national origin. EPA investigates these complaints in order to evaluate these allegations and determine if a violation of Title VI may have occurred. These investigations may involve performing an analysis of potential exposure and risk, and may also involve evaluation of methods for eliminating or reducing an impact.

EPA's Title VI *Draft Revised Investigation Guidance*[1] suggests that such assessments involve several discrete steps. These include determining the scope of an investigation, evaluating the potential for adverse impacts to occur in any affected population, and, if a possible adverse impact is identified, assessing disparity in impacts and in the composition of affected populations.

Administrative complaint number 16R-99-R9 (hereafter referred to as the Angelita C. complaint) was filed in June 1999 on behalf of children and parents who attend six California public schools with a high percentage of Hispanic students, by the California Rural Legal Assistance, Inc., Center on Race, Poverty & the Environment, California Rural Legal Assistance Foundation and Farmworker Justice Fund, Inc. The listed schools are:

- Rio Plaza Elementary School, unincorporated Oxnard, Ventura County, California
- Rio Mesa High School, unincorporated Oxnard, Ventura County, California
- Pajaro Middle School, Pajaro, Monterey County, California
- Barton Elementary School, Salinas, Monterey County, California
- Macquiddy Elementary School, Watsonville, Santa Cruz County, California
- Ohlone Elementary School, Pajaro, Monterey County, California

This assessment evaluates one of the complaint's several allegations, which pertains to possible adverse impacts to the complainants due to exposure to methyl bromide use in California, particularly in the vicinity of schools. The allegations contained in the complaint focus on the exposure of children at schools to the pesticide methyl bromide. The complaint notes that this chemical is extensively used to treat soil in unplanted fields, nursery soil, structural fumigation, and commodity fumigation. The complaint alleges that the re-registration of methyl bromide by the California EPA Department of Pesticide Regulation (CDPR) in 1999 had an adverse disparate impact on people of color, primarily Latino school children

As part of the technical support for conducting an investigation of this complaint, ICF developed methods to estimate methyl bromide concentrations at California public schools from

[1] *Federal Register*, June 27, 2000, Vol. 65, No. 124, pp. 39649-39701.

nearby agricultural applications. In support of that effort we have developed a regression model that can be applied to each such receptor school location in California (Cohen et al. 2011). Key input data for the model include methyl bromide use data, adjusted for several factors likely to affect the relationship between methyl bromide soil application and the corresponding ambient concentrations at the school location. These adjustment factors include the approximate distance of an application location from the school receptor location and the angle between the application and the school relative to the wind direction.

This document describes the development of several data bases used to support the application of the regression model to estimate methyl bromide concentrations at all California public schools, as well as the location and characteristics of receptor schools. This includes the composition of student populations associated with modeled concentrations and exposure averaging periods of interest.

2 BACKGROUND AND PURPOSE

Methyl bromide (MeBr), as a "restricted use" pesticide under California law, requires that an application notice of intent must be submitted to the CDPR through the office of a County Agricultural Commissioner. Once the application has taken place, the amount of chemical actually applied is reported to the CDPR. Growers are responsible for reporting the usage and treatment area to CDPR, and this requirement usually includes a hand-drawn paper map displaying the application field location.

The application records are entered into a database at the county level, and submitted to CDPR for maintenance in a statewide Pesticide Use Reporting (PUR) system. The state tracks pesticide usage in the database at the spatial resolution of Meridian Township Range Sections (MTRSs). An MTRS grid cell is typically one square mile (640 acres), although some are irregularly shaped. In addition to the MTRS, the amount of chemical and pesticide product applied, and the date of application, each usage record includes the number of acres treated as well as the number of acres planted in the field being treated

As described in the related model development document (Cohen et al. 2011), CDPR has previously developed models to predict MeBr concentrations based on usage in nearby MTRS areas. In order to apply the regression model to estimate both the potential exposure of specific schools with accuracy acceptable for this study, it was determined that MTRS-level spatial resolution for MeBr usage needed improvement if feasible. Indeed, the CDPR has noted that improving the spatial resolution of pesticide applications from the current square mile section to the operational level of an actual field site has been shown to significantly improve the ability to responsibly regulate pesticide use at the appropriate operational scale (Neal 2001 - slide 12). Furthermore, CDPR has noted that using section-based data has limitations in proximity analysis (Neal 2001 - slide 13), while field-based data greatly improve spatial accuracy (Neal 2001 - slide 15), as it pertains to knowing how much and where pesticide application is occurring. Previous studies (Ward et al. 2000; Rull and Ritz 2003) have also explored the use of more general land use data such as farmland identified through remote sensing to improve estimates of possible population exposures to pesticides.

Two major approaches were used in this analysis to develop more spatially refined usage estimates, consisting of the use of a set of smaller grid cells, combined with allocating the usage to these cells from the recorded MTRS based on the proportional amount of farmland in each cell. For this study, a geographic information system (GIS) was used to reapportion pesticide usage to sub-MTRS-section grid cells that are likely to be used for agricultural activities, i.e., that contain identified farmland. That is, reported MeBr usage in an MTRS was reapportioned among constituent grid cells, in proportion to the amount of farmland located in the grid cells. Implementation of this reallocation required development of several data bases described in this report.

The PUR database provided the initial location and other usage information for MeBr agricultural applications. The data were retrieved for the study time period (1995-2001), and reviewed for accuracy and completeness using methods similar to those developed by CDPR. The state's MTRS boundaries were initially used to evaluate the locations referenced in the PUR database where applications were reported to have occurred.

This study also involved development of an enhanced state-wide data base of California farmland locations that was created from multiple separate spatial data bases using GIS (see box). This data base is in the form of an Environmental Systems Research Institute, Inc. (ESRI) shape file, which allows the farmland areas to be viewed on a map.[2] For selected high-use counties, additional farmland data at the level of specific fields which were the locations of MeBr applications were also obtained and used.

Another data base developed for this study contains location information for all California public schools, as well as demographic information on the student populations. Development of this data base entailed merging data from two other data sets and an intensive quality assurance process for the school location coordinates.

A meteorology data base was also developed for this study containing data on wind speed, wind direction, and temperature measured at a number of station sites locations throughout California.

What is GIS? Geographical Information Systems (GIS) are computer hardware and software used for storage, retrieval, mapping, and analysis of geographic data.

Another key data base developed for this study was a set of tables generated using GIS tools and spatial analysis for the purpose of estimating the approximate distance of an application from the school receptor location and the angle between the application and the school relative to the wind direction. Interim GIS data sets used to create these tables included the Meridian Township Range Section (MTRS) grid for California, a customized ¼ mile by ¼ mile grid, the farmland

[2] Specifically, a shape file stores nontopological geometry and attribute information for the spatial features in a data set. The geometry for a feature is stored as a shape comprising a set of vector coordinates. For more information, see *ESRI Shapefile Technical Description: An ESRI White Paper*, July 1998, last accessed January 9, 2004 at http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf.

data base, the school data base, the meteorology data base, and MeBr usage data reported at the MTRS level.

3 PESTICIDE USAGE DATA

This analysis employed pesticide use data from the PUR (Cal-EPA/CDPR 2002) system, with several enhancements to the usage data, including identification of missing or erroneous individual records, updated usage outlier identification and correction processes, and improved spatial allocation. The PUR database contains records for MeBr usage with locations specified according to a meridian/township/range/section (MTRS) format of approximately 1-mile-square areas.[3] The entire state of California is divided into MTRS areas, and each agricultural usage record includes the MTRS containing the field over which the pesticide was applied, as well as the date (final application day in cases of multi-day applications), commodity (crop), formulated pesticide product, pounds of each active ingredient chemical applied, acres planted, acres treated, site (field) ID, and the grower ID (Cal-EPA/CDPR 2000). These 1-mile-square areas are the most precise location values included in the state's Pesticide Usage Report (PUR) database. This study apportioned the usage data in each MTRS to smaller grid cells as described below to improve the spatial resolution.

This analysis used the PUR MeBr usage report data for 2000 and 2001 monitoring days, as well as previous days up to eight days earlier[4]. The Pesticide Action Network- North America (PANNA) (http://www.pesticideinfo.org) had previously processed the entire set of PUR data, and provided an extraction of the MeBr data to us (personal communications Jan 10, 2003, May 5, 2003, June 23, 2003, Oct 3 2003.) Their processing also included some error correction and creation of outlier flags, following the procedures developed by CDPR[5] Although we originally

[3] The MTRS format is part of the Public Land Survey System (PLSS). The PLSS is a nationwide survey that grids the land in each state into approximately 1-square-mile rectangular units called sections. Some areas of California were not surveyed when California became a state because of Spanish land grants. We used a version of the PLSS prepared by CDPR with the grid lines extended to cover any areas that were not surveyed (Neal, 2001 - slide 8).

[4] These data include three records marked as database outliers by CDPR in the PUR database, all of which occurred in the year 2001 near the Kern CRS site. They were flagged by CDPR because they had very high usage rates (at least 300 pounds per acre) compared to the median for that commodity and product in 2001 (0.5 pounds per acre). These records were included in this analysis, however, because the usage rates in pounds per acre were not unusual for the commodity and methyl bromide product when compared to earlier years. CDPR also included these values in their regression modeling.

[5] Known types of errors contained in the California Pesticide Use Reporting (PUR) database include inexplicably high or low application rates, inaccurate acreage counts, incorrect site ID numbers, errors in the percent of active ingredients in a particular pesticide product. .In the late 1990s, CDPR developed error-checking routines that flag probable errors when application rates are too high. These records are flagged as 'outliers' and these data points are removed when working with the data. While the number of records in error flagged by this method is approximately 0.4% of the total number of records, errors in total pounds of pesticides used are quite large -- approximately 2–5% of the total pounds of reported use in the state -- on the order of 4–10 million pounds each year. Further information is available on CDPR's methodology for flagging outliers (Cal-EPA/CDPR 2000 and Wilhoit 2002) as well as their review of data quality (Wilhoit, Zhang and Ross, 2001).

Two general types of errors exist in the PUR data. The first is related to chemical use, the second related to acreage reporting. Historically, most of the error correction analysis has focused on correcting chemical application rates, specifically by flagging chemical application rates that are above normal. For this reason, the reported pounds

obtained the processed data from this secondary source, we also checked the PANNA-derived data against a retrieval of the same data from CDPR using their California Pesticide Information Portal (CalPIP) at the website http://calpip.cdpr.ca.gov/main.cfm and found no differences other than the modifications discussed below that we had decided to make.

3.1 Duplicate Usage Records

The goal of this step in the exposure assessment was to find and remove all usage records suspected of being duplicates of other records, to avoid overestimating MeBr exposures. Our methodology was based on an initial review of 68 suspected duplicate records of 2000 pounds or more for the 1999 to 2001 period. Some 28 of these 68 records were determined not to be likely duplicates. Two of those 28 were confirmed to not be duplicates by the Tulare County Agricultural Commissioner, who informed us that they were likely a case where the grower applied the MeBr, using an untarped method, on the exact same field at half the usual application rate on one day and then a few days later applied the same amount over the same field. This approach allows the grower to significantly reduce the size of the CDPR required buffer zone warning area. (Unfortunately, the hard copies of these records have been destroyed, so that these circumstances could not be documented.) There were 38 records judged likely to be duplicates. Thirty of these 38 were confirmed to be duplicates by the appropriate county commissioners. The remaining 2 of the original 68 records had an uncertain status. Details of the approach are presented in Appendix A.

3.2 Usage Outliers

CDPR has developed procedures to identify and flag usage records which may be erroneous because they are outliers compared to typical usage rates for similar records.(Wilhoit 1998, 2002, 2005; Cal-EPA/CDPR 2000). CDPR used three application rate outlier criteria and defined a record to be an outlier if any of these three criteria were met. The criteria were

- More than 1000 pounds of MeBr used per acre
- More than 50 times the median product used per unit treated for the same commodity, product, and measurement unit, or more than the median plus 50 times the median absolute deviation (MAD)
- Neural network: a procedure that maps the summary statistics (e.g. mean, median) of the application rates for the current year, commodity, product, and measurement unit into an upper threshold value, based on a review by 12 CDPR scientists of “typical” distributions of application rates

---

of active ingredient is generally accepted as being an accurate number, within the constraints of a system that relies mainly on hand data entry. The most significant errors remaining in the PUR data are in the reported number of "acres planted" for some crops, and work on correcting these errors has begun within CDPR. However in the meantime, because acreage data in the PUR dataset still contains errors (some of them quite large), PANNA’s data processing techniques and error correction routines should be used. For a more complete description of the PANNA methodology, see: http://www.pesticideinfo.org/Docs/ref_PURCA.html#PANProcess.

There were 45,822 MeBr agricultural usage records in our database (from 1995 to 2001), of which 198 were removed as duplicate records as described above, leaving 45,624 unduplicated records. CDPR reported a total of 456 outliers (1.0 %), i.e., 456 records met at least one of their three outlier criteria. The CDPR analyses used comparison criteria (e.g. median values) which were based on each current year of data only. For some commodities which relatively infrequent usages, these criteria could vary somewhat from year to year, which may have been a result of the small number of such uses rather than a true measure of overall usage variation.

Generally, our preference was to modify as few records as reasonable, which almost always reduces them. Therefore, for this analysis we used a slightly different set of criteria to identify usage outliers as follows, which is believed to have identified erroneous records while avoiding "false positive" identifications of records that were still reasonable in the context of the usage history for that commodity, while unusual for a particular year,..

- More than 1000 pounds of MeBr used per acre
- More than the mean plus three standard deviation of the logarithm[6] of MeBr used per unit treated for the same commodity, product, and measurement unit[7] (e.g. pounds per acre).
- Neural network: a procedure that maps the summary statistics (e.g. mean, median) of the application rates for the current year, commodity, product, and measurement unit into an upper threshold value, based on a review by 12 CDPR scientists of "typical" distributions of application rates. CDPR had developed the thresholds based only on the data in the current year. But for this analysis this criterion was applied only if there was evidence of some stability in the year to year distributions, i.e., if the median application rates for the commodity, product, and measurement unit in the current and previous years were within a factor of 10, i.e. the ratio was between 0.1 and 10.

Out of the 45,624 non-duplicate agricultural records for 1995-2001, CDPR reported 456 (1.0 %) outlier records, i.e., records that met at least one of the three criteria. Our alternative approach produced a total of 329 (0.72 %) outlier records that met at least one of our three criteria. Of these 329 records, 207 were also outliers according to CDPR, but 122 of these 329 records were outliers according to our criteria but not according to CDPR's criteria, i.e., they met our Criterion 2 and not CDPR's.

As suggested by CDPR's initial treatment of suspected outliers, any records found to be outliers were adjusted by reducing the application rate to the median level.

Details of our approach and its comparison to CDPR's approach are presented in Appendix B.

[6] The logarithm of the application rate (pounds of product used per unit treated) was used for these calculations, since CDPR noted that the distribution of the application rate is skewed. This suggests that a log transformation will make the data more symmetric and closer to a normal distribution.

[7] In order to obtain sufficiently large sample sizes (i.e., at least 50 records) for making reasonably accurate approximations of the true mean and standard deviation, a tiered approach for grouping records across years, commodities, and products was adopted. See details in Appendix B.

3.3 Multi-day Applications

Generally, up to the year 2001, growers were expected to make a single report of the total usage dated at the final application day rather than reporting each daily application. The approach used for determining which records are likely to refer to a multi-day application and for allocating the reported MeBr usage for such records across multiple days over the time period of study. was based on conversations about standard practices with Paul Niday of Tri-Cal applicators (personal communication October 30, 2003), one of the largest MeBr applicators in the State, and three county agricultural commissioner officers in three counties, Ventura, Tulare and Riverside (personal communications of October 1, 2003, October 2, 2003, and October 8, 2003), where large MeBr applications have historically occurred.

The approach is divided into two time periods: those applications which occurred prior to the revised State MeBr application regulations which became effective on 14 January 2001, and the applications following these new procedures. The new regulations decreased the maximum allowable application rate and increased the minimum waiting time between applications at a single location. These regulations (and more recent updates) are fully described in the California Code of Regulations (Cal. Code Regs., Tit. 3, §6447)..

When our approach indicated that an application was conducted over multiple days, we apportioned the usage across the days under the assumption that 40 acres could be treated per day prior to January 14 2001, and 30 acres per day thereafter. For the application records whose actual notification reports were obtained from county agricultural commissioner offices, the actual days of application were included, since the reports included this information.

Appendix C describes the approach for identifying and temporally allocating multi-day usages in more detail.

3.4 The Projection and MTRS GIS File

Pesticide usage information is reported in California at the spatial resolution of Meridian Township Range Sections (MTRSs). MTRSs are mapped land areas of approximately one square mile (see text box). The MTRS is used by the California Department of Pesticide Regulation (CDPR) to track the amount of pesticides applied in locations around the state.

For this study, a California-wide MTRS GIS file was obtained from PANNA. The original GIS file was developed by CDPR from its PLSNet data set. PLSNet is an MTRS database developed from existing surveyed section lines to provide a complete coverage of California for pesticide use reporting. Data are distributed by county as PLSxx.shp files, where xx is the county number (e.g., 01 = Alameda).[8] The PLS layer in the GIS file is a polygon coverage depicting the township, range and sections contained in the PLS grid for the State of California.

California uses three base line/meridian pairs:

[8] See additional detail at http://www.cdpr.ca.gov/docs/county/pumpdvlp/cacconf/conf2000/howto.htm, last accessed January 20, 2004.

1. Humboldt (northwestern corner of the state), or HB&M;
2. Mt. Diablo (most of northern California), or MDB&M; and
3. San Bernardino (southern California), or SBB&M.

The township and range values of the PLSnet data are combined in the redefined item TOWN-RANGE to facilitate dissolve (the merging of adjacent polygons or lines that have the same value for a specified item) and other functions.

For pesticide use reporting purposes, CDPR expanded the existing PLS data and developed "sections" in previously non-surveyed areas of California to create a complete statewide coverage (Neal 2001 - slide 8). Most of the areas in the State that are not gridded into sections (as described in the box above) involve Spanish and Mexican land grant areas that were honored by the U.S. when California became a State and were subsequently excluded from the section survey process. Names of land grants in the GIS file may be found under the item LAND GRANT. Areas not sectioned previously because of difficulties in surveying wetlands and mountainous terrain are identified in the item COMMENTS as "wetlands" or "not sectioned."(Nordstrom 1996).

The **Meridian Township Range Section (MTRS)** system is based on the original Public Land Survey (PLS) system that began in the late 1700's to create parcels of land primarily in the Midwest and West suitable for disposal by the Government. The PLS uses various *initial points* through which an east-west line, or *base line*, and a north-south line, or *principal meridian*, are designated. From this initial point and oriented along the baseline and meridian, approximately six-by-six mile squares called townships typically are laid out. Each township is identified by both a T number according to how far north or south it is of the designated baseline and a range, or R, number according to how far east or west it is of the designated meridian. For example, T32N, R18E is the township that is at the intersection of the 32nd row of townships north of the designated baseline and 18th column (or range) east of the designated meridian.

The townships in turn typically are divided into 36 roughly one square mile sections, or MTRSs, numbered starting in the northeast corner of the section. The numbering increments moving west at first and then "boustrophedonically", or "as the oxen plows and turns", which means that the numbering wraps around in an "s" shape from north to south.

Most lands have been surveyed under PLS. Some lands were excluded originally because of difficulties in surveying wetlands and mountainous terrain, or the lands were national installations such as military reservations and national parks, or they were areas such as land grants that had already passed to private ownership prior to subdivision.

For this study, the GIS data file obtained from PANNA was projected to a specialized Albers Equal Area projection, similar to the Teale Albers Equal Area projection used by California State.[9] Like Teale, the specialized Albers projection assigns a central meridian of 120 degrees, which runs approximately through the longitudinal center of the State, and which is useful in assuring accurate area calculations. Additionally, the first and second standard parallels are defined by the northern- and southern-most latitudes of the State. The projection parameters are presented in Table 1.

[9] As described in more detail at a CDPR permit/use mapping developers group website, http://www.cdpr.ca.gov/docs/county/pumpdvlp/devgrp/prjctdta/pres0401.htm (last accessed January 21, 2004), CDPR and a number of other state agencies use a specific projection called the Teale Albers Equal Area projection. The parameters for this projection are: 1st Standard Parallel: 34 00 00; 2nd Standard Parallel: 40 30 00; Central Meridian: -120 00 00; Latitude of origin: 00 00 00; False easting (meters): 0; Datum: NAD27; Spheroid: Clarke 1866; and Units: meters.

Table 1. Projection Parameters

Parameter	Value
Projection	ALBERS
Datum	NAD83
Zunits	NO
Units	FEET
Spheroid	GRS1980
Xshift	0.0000000000
Yshift	0.0000000000
1st standard parallel	32 0 0.000
2nd standard parallel	42 0 0.000
Central meridian	-120 0 0.000
Latitude of projection's origin	30 0 0.000
False easting (meters)	0.00000
False northing (meters)	0.00000

3.5 Temporal Allocation of Usage Data

The computerized data contains only a single date for each MeBr application. Based on current reporting regulations, multi-day applications are reported as a single record, dated by the last application day, in the PUR. In order to identify such records so that the usage could be appropriately apportioned across multiple days for the daily average regression model calibration, we investigated cases where the number of acres treated was large enough to suggest the strong possibility a multi-day application.

Where available, we obtained copies of the paper records (Notice of Intent and Pesticide Usage Report) for all the large applications (above 30 acres) with more than 100 pounds of MeBr per acre treated. These paper copies include the actual application days and, in most cases, the acres treated on each day. We reapportioned the usage across the individual days proportionally to the acres treated on that day.

However, because these reports were not available for all applications, another approach was used to supplement the site-specific information. When large amounts of usage were reported which had been found to be typically associated with multi-day applications, the usage was assumed to have been spread over two or more days, as a function of the total amount applied and the acres treated information.

4 MASTER CALIFORNIA FARMLAND DATA BASE

Many pesticides are more likely to be associated with farmed than non-farmed land. Farm-related chemical usage are recorded in California at only the meridian-township-range-section

(MTRS) level,[10] which thus associates the applications with a wider area than they are in fact associated (i.e., specific fields). With the use of farmland data, however, these can be reapportioned to specific sub-MTRS areas closer to where they are more likely to exist. Unfortunately, such finer spatial information detailing exactly where and when a farm-related activity occurs is not collected consistently or in an easy-to-use manner by California counties at this time. Furthermore, no single source of sufficiently detailed farmland data for the State of California could be found. To develop such a database, therefore, it was necessary to composite multiple sources of data to obtain full coverage of the State.

As data sets were obtained for this development of a detailed California farmland database, a preferred sequence of data use was established based on the quality of the input data sets. An effort was made to collect data most closely representing contemporary agricultural fields at which pesticides are most likely to be applied. For example, efforts to exclude pasture land, often included in agricultural data sets, was made since MeBr applications included in this assessment do not apply to pasture land. Also, the National Land Cover Data (NLCD) set was excluded because it was found to be unreliable with regard to its classification of single and clustered agricultural pixels (see section 4.1.3 below).

This section steps through the original data review, integration, and processing and the final farmland data set characteristics. Six sources of data were identified for potential use in the development of this composite of California farmland:

(1) California Department of Conservation, Division of Land Resource Protection, Farmland Mapping and Monitoring Program (FMMP),

(2) California Department of Water Resources, Division of Planning and Local Assistance Land Use Survey Data (LUSD),

(3) National Land Cover Data (NLCD),

(4) California Department of Forestry and Fire Protection, Fire and Resource Assessment Program (FRAP),

(5) Santa Cruz County Parcel Data, and

(6) Monterey Ranch Data.

These data sets were initially identified in part because they were readily available in GIS format and were therefore easy to integrate into the reapportioning effort. Also, while each only covered a subset of the area of interest, together they formed a complete coverage of farmland in California.

[10] An MTRS section, or Meridian Township Range Section, is a defined area within the bounds of the surveyed Public Land System grid, or within the expansion of that grid to include non-surveyed areas. It is typically a square land parcel that is one square mile in area.

The above data sets were specifically reviewed for how accurately they spatially portrayed current farmland distribution. Aerial photography was used to verify the spatial and temporal accuracy of some of these data sets, as was easily available, and particularly in the determination of whether or not to include the NLCD satellite imagery based classification. Four data sets—FMMP, LUSD, FRAP, and Santa Cruz—were ultimately selected for use.

The integration of the four farmland data sets was conducted in the ESRI ArcGIS environment using a series of Arc Macro Language (AML) programs.[11] The AML programs handled all aspects of the integration including data conversions, projections, unions, and error handling. These programs also addressed both the basic overlay hierarchy, which allowed the credibility of the data source to dictate the order in which it is assembled (so that the more authoritative layers always overwrote the less authoritative layers where both were available), and the artifact handling, which is a key characteristic of how the final farmland data set is composed. Specific details on the farmland data processing are provided below, including the review conducted on existing farmland resources, the integration of multiple data sources into a single database, and the artifact and small polygon removal processing. An additional section describes the characteristics of the final collection of generalized farmland data used in this assessment.

4.1 Review of Existing Farmland Resources

4.1.1 California Department of Conservation, Division of Land Resource Protection, Farmland Mapping and Monitoring Program (FMMP)

As mentioned above, six sources were reviewed for inclusion in the composite farmland data set. The source deemed the most reliable was the California Department of Conservation, Division of Land Resource Protection, Farmland Mapping and Monitoring Program (FMMP) data.[12] The FMMP data used in this study represent the 1998-2000 timeframe and cover most of California. This data source was developed originally for the purpose of understanding impacts on California's agricultural lands, and it includes a ranking system of agricultural purposes. Also known as Important Farmland data, the FMMP farm data are compiled using U.S. Department of Agriculture, Natural Resources Conservation Service (USDA-NRCS) Soil Surveys, current land use information, and the results of a public review process.

The minimum mapping unit is 10 acres, although a small set of polygons smaller than 10 acres were received with the original shape files. The data are updated every two years using a combination of aerial photography taken at various scales (mostly 1:130,000 scale color infrared), field reconnaissance, and a review of 1:24,000 scale U.S. Geological Survey (USGS) topographic quadrangles. The original data are assembled at the county level and these are individually updated and/or field-checked using varying methodologies, data sources, and ground-truthing. The data are available as ESRI Arc shape files on a county-by-county basis in the "DOWNLOAD FMMP GIS DATA" section of the FMMP Statistics and Reports website.[13]

[11] AML is the legacy programming language of ArcInfo. The program series used to develop the farmland data is provided in Appendix E.
[12] Last accessed on January 8, 2004 at http://www.consrv.ca.gov/dlrp/fmmp.
[13] Last accessed on January 9, 2004 at http://www.consrv.ca.gov/DLRP/fmmp/Pages/Index.aspx.

The methodology, source of aerial imagery, source data, scale, film type, and information about coverage gaps is available at this site.

Land use in the FMMP is divided into 14 categories of which the following 11 categories were selected for this composite (see above reference for more detail and Appendix D for full metadata):

(1) Prime Farmland (P) - Irrigated land with the best combination of physical and chemical features able to sustain long-term production of agricultural crops. This land has the soil quality, growing season, and moisture supply needed to produce sustained high yields. Land must have been used for production of irrigated crops at some time during the four years prior to the mapping date.

(2) Farmland of Statewide Importance (S) - Irrigated land similar to Prime Farmland that has a good combination of physical and chemical characteristics for the production of agricultural crops. This land has minor shortcomings, such as greater slopes or less ability to store soil moisture than Prime Farmland. Land must have been used for production of irrigated crops at some time during the four years prior to the mapping date.

(3) Unique Farmland (U) - Lesser quality soils used for the production of the state's leading agricultural crops. This land is usually irrigated, but may include non-irrigated orchards or vineyards as found in some climatic zones in California. Land must have been cropped at some time during the four years prior to the mapping date.

(4) Farmland of Local Importance (L and LP) - Land of importance to the local agricultural economy as determined by each county's board of supervisors and a local advisory committee. See above reference for either "A Guide to the Farmland Mapping and Monitoring Program", or the latest copy of the "Farmland Conversion Report" for each county's definition of Farmland of Local Importance.

(5) Grazing Land (G) - Land on which the existing vegetation is suited to the grazing of livestock. This category is used only in California and was developed in cooperation with the California Cattlemen's Association, University of California Cooperative Extension, and other groups interested in the extent of grazing activities. The minimum mapping unit for Grazing Land is 40 acres.

(6) Urban and Built Up Land (D) - Land occupied by structures with a building density of at least 1 unit to 1.5 acres, or approximately 6 structures to a 10-acre parcel.

(7) Other Land (X) - Land that does not meet the criteria of any other category. Typical uses include low-density rural development, heavily forested land, mined land, or government land with restrictions on use.

(8) Water (W) - Water areas with an extent of at least 40 acres.

(9) Irrigated Farmland (I) - Cropped land with a developed irrigation water supply that is dependable and of adequate quality. Land must have been used for production of irrigated crops at some time during the four years prior to the mapping date.

(10) Non-irrigated Farmland (N) - Land on which agricultural commodities are produced on a continuing or cyclic basis utilizing stored soil moisture.

(11) Area Not Mapped (Z) – Areas that fall outside of the NRCS soil survey and which are mapped by the FMMP.

4.1.2 California Department of Water Resources, Division of Planning and Local Assistance Land Use Survey Data (LUSD)

The second most reliable data source identified for farmland polygons was LUSD from the California Department of Water Resources,.[14] The data were gathered using aerial photography and extensive field visits, the land use boundaries and attributes were digitized, and the resultant data went through standard quality control procedures before finalizing. The LUSD is updated using aerial photography of varying scales, sources and flight dates – as well as incorporating methodologies using fieldwork and USGS topographic quadrangles. Each county is individually updated and field-checked using varying methodologies, data sources, and field investigators. Information regarding the county-by-county updates can be found on the Department of Water Resources website. The metadata file (available for each county) explains updates in detail including source, data development processes, accuracy as well as contact information. These data represent farmland for the 1976-2003 timeframe.

Polygons extracted for the farmland dataset are those associated with the following. (LUSD Standard Land Use codes are in parentheses.)

- Grain and hay crops (G)
- Rice (R)
- Field crops (F)
- Truck, nursery and berry crops (T)
- Deciduous fruits and nuts (D)
- Citrus and subtropical fruit (C)
- Vineyards (V)
- Idle crop land (I), and
- Turf farms (P7)

See Appendix E for full metadata.

[14] Last accessed on January 9, 2004 at http://www.water.ca.gov/landwateruse/lusrvymain.cfm.

4.1.3 U.S. Department of Interior, United States Geological Survey, National Land Use Cover (NLCD)

Derived from the early to mid-1990s using Landsat Thematic Mapper satellite data, the National Land Cover Data (NLCD) is a 21-class land cover classification scheme applied consistently over the United States.[15] The spatial resolution of the data is 30 meters and mapped in the Albers Conic Equal Area projection, NAD 83. The NLCD are provided on a state-by-state basis. The state data sets were cut out from larger "regional" data sets that are mosaics of Landsat TM scenes. The TM multi-band mosaics were processed using an unsupervised clustering algorithm. Both leaf-off and leaf-on data sets were analyzed. The resulting clusters were then labeled using aerial photography and ground observations. Clusters that represented more than one land cover category were also identified and, using various ancillary data sets, models were developed to split the confused clusters into the correct land cover categories. These can be found on the Geographic Data Download section of the United States Geologic Survey (USGS) website referenced above. The State of California is divided into a north and south section. Metadata information, a readme file, accuracy reports and other information are available from the above website. The readme file contains information regarding which Landsat scenes (Path/Row, Date) were included in the analysis.

Ultimately, however, these data were not integrated into the final composite due to the lack of reliability in correctly classifying single and clustered pixels as farmland (see Appendix F). The data are generally not suited to fine-scale analysis. Additionally, the identification of falsely-attributed agricultural land in urban centers rendered this data set inappropriate for this compilation. (http://landcover.usgs.gov)

4.1.4 The California Department of Forestry and Fire Protection, Fire and Resource Assessment Program (FRAP)

FRAP is a compilation of the "best available" land cover data into a single data layer, to support the various analyses required for the 2002 Forest and Range Assessment.[16] Typically the most current and detailed data were collected for various regions of the State. Decision rules were developed that controlled which layers were given priority in areas of overlap. Data sources had unique scales and resolution, and therefore the final FRAP product was initially standardized and developed as a 30 meter resolution ESRI Arc GRID. This was later resampled for distribution to one hectare (10,000 sq. meter, or 100 meter length) resolution when the file sizes of the 30 meter GRIDs exceeded processing limitations of the FRAP program.

The original January 2002 data used to support the assessment is also available from the FRAP website – however, the most recent version available at the time, from October 2002, was used in this study, which incorporates newer data for the Mojave and NE Colorado Desert areas. Agricultural polygons drawn for this study are sourced from/or verified by the following six data sets (see Appendix G for full metadata):

[15] Last accessed January 9, 2004 at http://edc.usgs.gov/geodata/.
[16] http://frap.cdf.ca.gov/projects/frap_veg/index.asp

(1) FMMP,
(2) Department of Fish and Game 1999 wetlands data,
(3) California Department of Forestry and Fire Protection/USDA Forest Service CALVEG (1994-1997),
(4) CDF Hardwood Rangelands (1990),
(5) Yosemite National Park (1934) vegetation data, and
(6) 1996 GAP analysis project UC Santa Barbara.

The minimum mapping unit for FRAP is 2.47 acres (one hectare).Only polygons labeled as "agricultural" were used in this composite. Although a primary source for FRAP was the FMMP, FRAP rejected some of FMMP's agricultural categories, such as the "Locally important" class because it was not consistently applied across the state. Additionally, FRAP only included FMMP classes "Prime," "State," "Unique," and "Developed". In addition, agriculture from several other data sources contributes to the statewide product. Finally, non-agricultural areas from data sources given a higher standing in the overlay process (lakes, urban, riparian) overwrote some areas mapped by FMMP as agriculture.

4.1.5 Santa Cruz County Farm Parcels

Although few counties in California possess GIS data for individual farm parcels, such information was available for Santa Cruz County from the Agricultural Commissioner's office. Santa Cruz continually updates their digital parcel database by digitizing the polygons from the county one-foot aerial photography with every new permit or field planting re-configuration. The parcel polygons are tagged with the PMTSITEID, which are the permit number concatenated with the Site ID. This unique parcel identification system enables linkage back to all pesticide usage reported for that field. However, the major limitation to this data set is that as new permits are issued, the PMTSITEIDs are deleted and reassigned, rendering a link to past usage for a particular parcel impossible. The data set acquired for this data compilation represented parcel polygons from 2002 that included some unchanged parcels whose permits dated back to 2000 and 2001.

When compared to the farmlands from the FMMP database for Santa Cruz County, the Santa Cruz data were, for the most part, spatially aligned with the FMMP data. However, the Santa Cruz information was more accurate at delineating between actual farm parcel and farm-related infrastructure, water bodies, roads and other aspects of the landscape that went beyond pure farmed lands. FMMP information for Santa Cruz was completely replaced by Santa Cruz County information. However, it is important to note that the Santa Cruz data obtained for this data compilation represents a single time-slice (2000-2002) and is complete for lands specific to only this one pesticide.

Minor editing of the individual parcels was required for the purpose of removing slivers so that the union of the Santa Cruz data to the other data sources for the areas outside of Santa Cruz would proceed smoothly. Additionally, a polygon approximating the county boundary but fit to the actual edges of polygons received and approximately the river centerline, was used to both cut a hole in the larger FMMP data, as well as to fit the better Santa Cruz data into the composite. The primary area of concern was the border between Santa Cruz and Monterey

counties. For the purposes of reapportioning reported pesticide usage to a finer resolution grid, identification of where the Monterey farmland stopped and where Santa Cruz farmland began was required. Since the Santa Cruz farmland was the most accurate digital data available and was digitized from one-foot aerial photography, it was used to establish the boundary between the two counties.

4.1.6 Monterey Ranch Data

Monterey County contains much farmed land and efforts were made to research a better source of farm parcel boundaries for the purposes of this data development effort. Indeed, the Monterey County ranch data was found but in the end, not used other than for quality assurance purposes. The ranch data are developed from an ownership parcel database. The data therefore do not follow along farmed land boundaries, but rather along boundaries between different owners. In effect, houses, rivers, and other non-farmed land appear to be indistinguishable from the farmed lands around them (see Figure 1, where FMMP lands are in green and ranchero lands are outlined in black). Additionally, the "ranchero" system was originally an ownership-based system, and therefore ranchero lands that are now experiencing non-agricultural usage, such as Salinas Airport, are still parcels included in the ranch data set.

Figure 1. Ranchero and FMMP data

Ranchero = black line
FMMP = green areas

Furthermore, upon spatially overlaying the ranchero polygons with digitized fields derived from pesticide applications, it was determined that the names of the ranches in the digital ranchero dataset appear to be historic, rather than reflecting the current ownership or grower name information. This further impedes the chances of linking the actual farmed areas to specific usage information. And even though permit information is contained in the ranch data, this information only indicates whether or not the owner has filed notice for a specific application. While this is useful information, it does not indicate when the application permit was granted, which acreage received any pesticide application, whether pesticides were actually applied, and whether pesticides may previously have been applied to currently un-permitted parcels.

Given the extent to which the Monterey County ranch boundaries generally track the FMMP data, and given the lack of complete information from Monterey, it was decided to use the FMMP data for the general farmland dataset.

4.2 Integration of Multiple Data Sources

The FMMP data set from the California Department of Conservation was used as the main source of data for this study. The categories of P, S, U, L, and LP (as described in Section 4.1.1) were selected to define "Farmland". To define "Not Farmland", G, D, X, and W were used. Z

was used to determine areas that were not mapped. The polygons that fell into each category were combined (dissolved), creating a dataset with three categories: "Farmland"; "Not Farmland"; and "Not Mapped". These category selections and dissolves were applied to each available county's worth of data, and finally merged into a single database.

Since there was a significant amount of the state that the FMMP data had not mapped, other data sources were needed as mentioned above. The LUSD from the California Department of Water Resources was deemed the next most reliable data source. The LUSD was similarly reclassed into three classes with the following components (see Section 4.1.2): "Farmland" included G, R, F, T, D, C, V and I; "Not Farmland" included N, U, S, and P; and "Not Mapped" was comprised of NS, E, and Z. Like FMMP, the LUSD data are divided on a county basis (or county equivalent). The above procedure was done for all available counties. The results were dissolved and all counties merged together and finally combined with the FMMP data where the FMMP class was "Not Mapped".

Still, gaps existed after these two data sources were joined. For areas where both the FMMP data and the LUSD data were "Not Mapped," data from FRAP was used. This raster (or grid) dataset was converted to a vector (or polygon) dataset. Only the agriculture category ("02") was selected for integration. FRAP data cover the entirety of the state, so it was combined with the previous two datasets to fill in where those two were collectively "Not Mapped". Due to the resampling of the data to one hectare grid cells, even where FRAP and the other two sources essentially agreed in terms of the spatial location of the farmland, "stair-stepped" artifacts were generated by the creation of the composite. The bulk of these artifacts were removed using an automated approach described in the next section.

Finally, the Santa Cruz County Farm Parcels were selected for that county, and replaced any farmland data allocated to that county from the other data sets.

4.3 Removal of Small Polygons and Artifact Polygons

Two aspects of the process of creating the composite farmland data set are described in this section. The first is the removal of all polygons smaller than two acres from three of the four original data sources, prior to their integration into the composite. The second is the removal of artifact polygons from the final data composite data set, which were created as an unintended byproduct of the joining process. Both steps are intended to yield a more reliable and smoother rendering of the farmland polygons.

The decision to remove the sub-two acre polygons was based on a review of the underlying minimum mapping units for each data source. They are as follows:

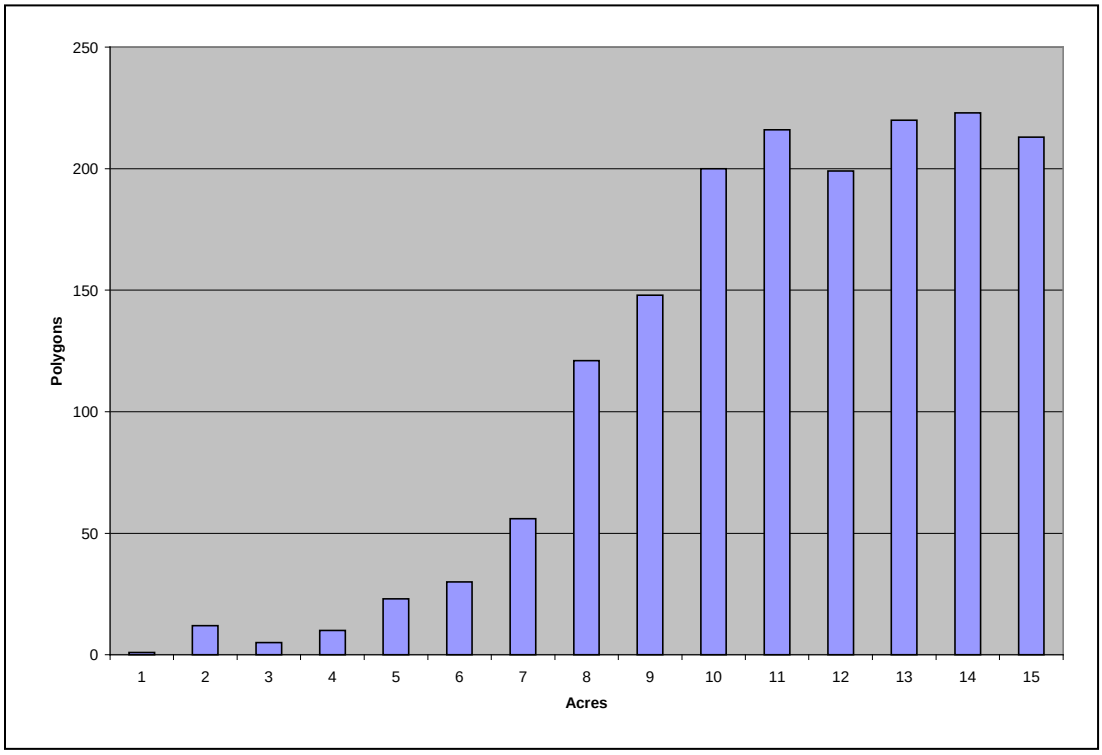
FMMP	10 acres
LUSD	2 acres[17]

[17] Although the precise minimum mapping unit for LUSD is not reported in the available documentation, correspondence with Ed Morris, California Department of Water resources, emorris@water.ca.gov, indicates that an approximate minimum mapping unit is in the area of 2 acres for field verification. FRAP documentation, which

FRAP 2.47 acres
SC no minimum

Despite the reported minimum mapping units, polygons below these sizes were found in the original data sets. As described above, the desired farm-related categories were selected from the sources, and then dissolved into the three categories (or one category in the case of FRAP). Histograms showing the frequency of small polygons (15 acres or less) after internal farmland boundaries were dissolved, are seen in Figures 2, 3, and 4.(Note that the label on each bar of each histogram specifies the upper end of a 1-unit size range.)

Figure 2a. Distribution of small polygons in the FMMP data set, farmland only.



references LUSD polygons, cites the LUSD minimum mapping unit as 1 acre based on the use of 1:24,000 scale rubber-sheeted aerial photographs.

Figure 2b. Distribution of small polygons in the FMMP data set, including "not farmland"

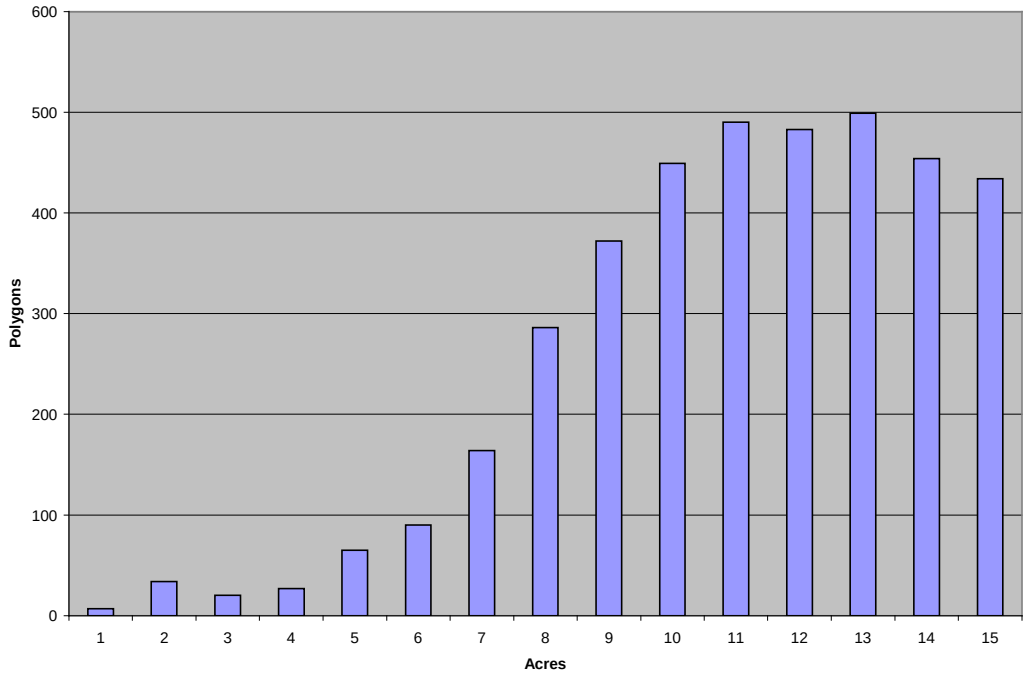


Figure 3a. Distribution of small polygons in the LUSD data set, farmland only

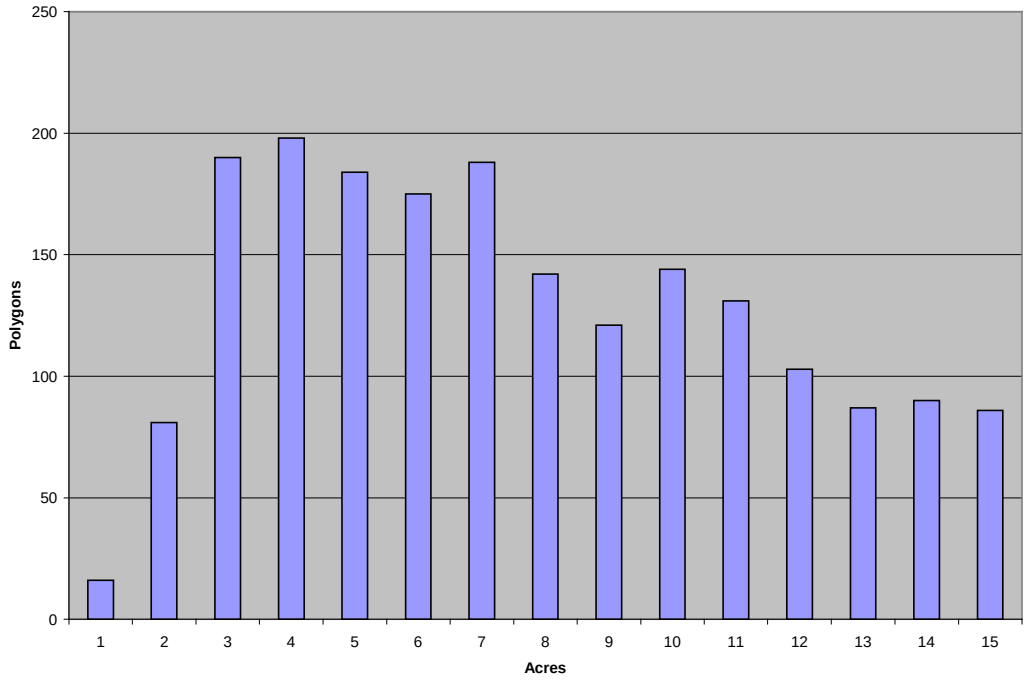


Figure 3b. Distribution of small polygons in the LUSD data set, including "not farmland"

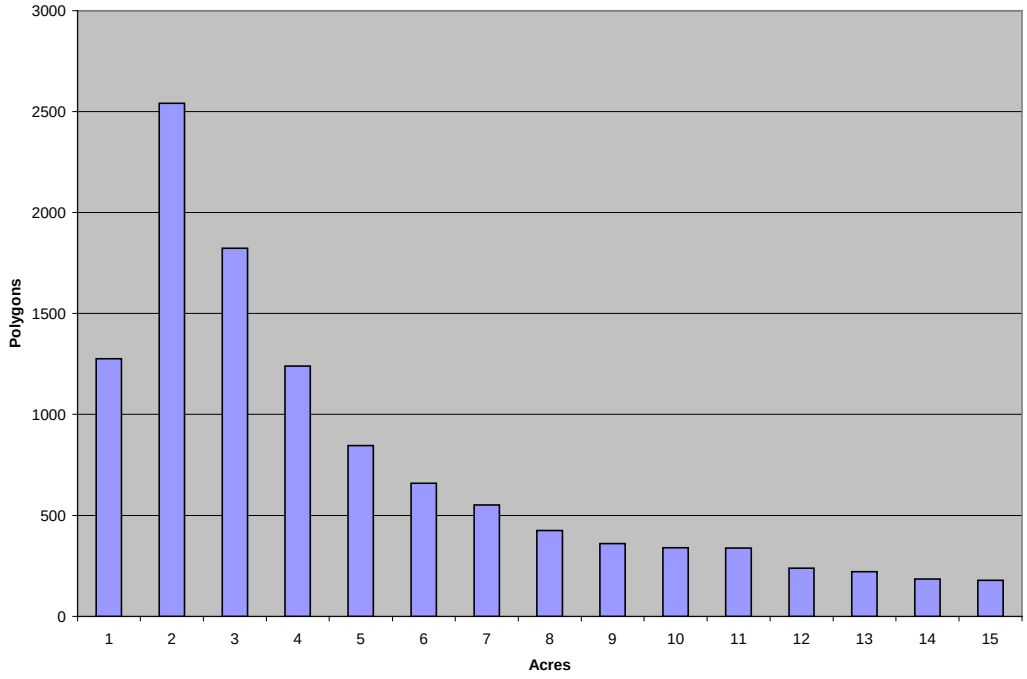
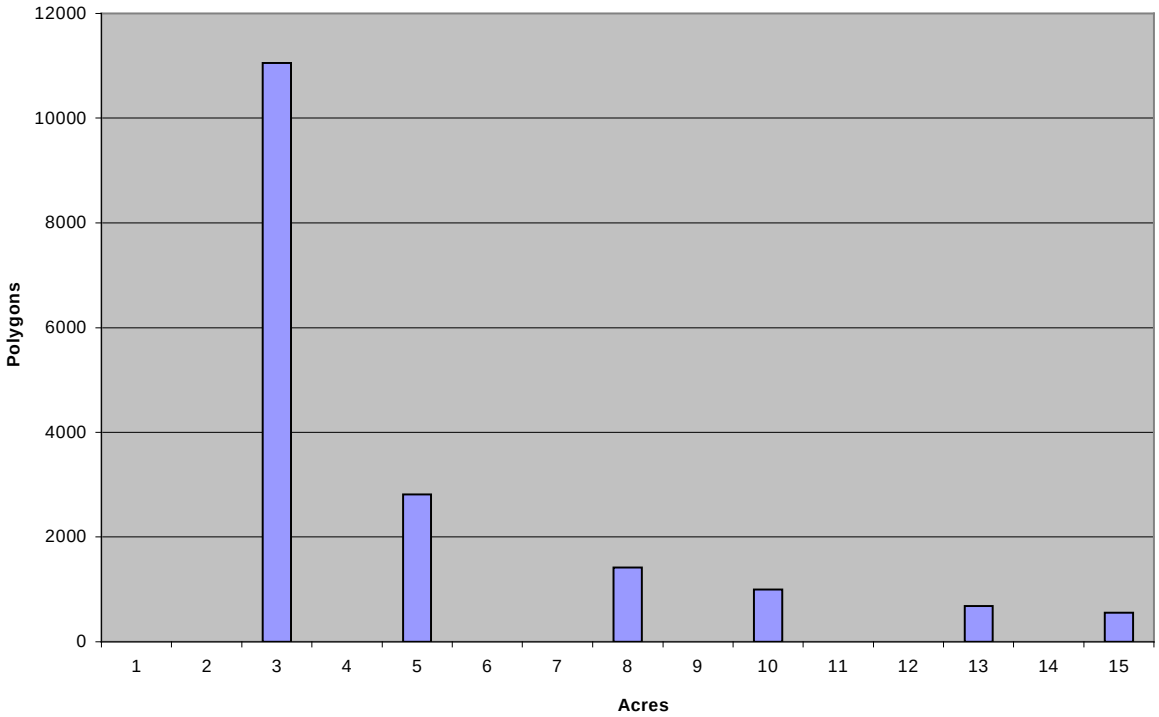


Figure 4. Distribution of small polygons in the FRAP data set, farmland only



Together, the information regarding the minimum mapping units and the small polygon distribution led to the decision to remove all polygons in the original data sources, from the post-dissolve version of the data that were smaller than 2 acres. Polygons this small seemed unreliable given the reported methodologies and data characteristics. We predicted and coded for the removal of three types of sub-two acre polygons in the data. These are:

(1) Sub-two acre polygons that are entirely surrounded by a polygon with a different category. For example, see Figure 5 for a "Not Farmland" polygon (marked with a blue border) surrounded entirely by a "Farmland" polygon.

(2) Sub-two acre polygons that sit on the edge of a polygon with a different category. For example, a small polygon of "Farmland" adjacent on one side to a large area of "Not Farmland". (None of these were actually present.)

(3) Sub-two acre polygons that are surrounded by "Not Mapped" areas. For example, see Figure 6 for a "Not Farmland" polygon (marked with a blue border) that is surrounded by a "Not Mapped" area (in this example, the white area is "Not Mapped").

Figure 5. Example of "Not Farmland" polygon surrounded by a "Farmland" polygon

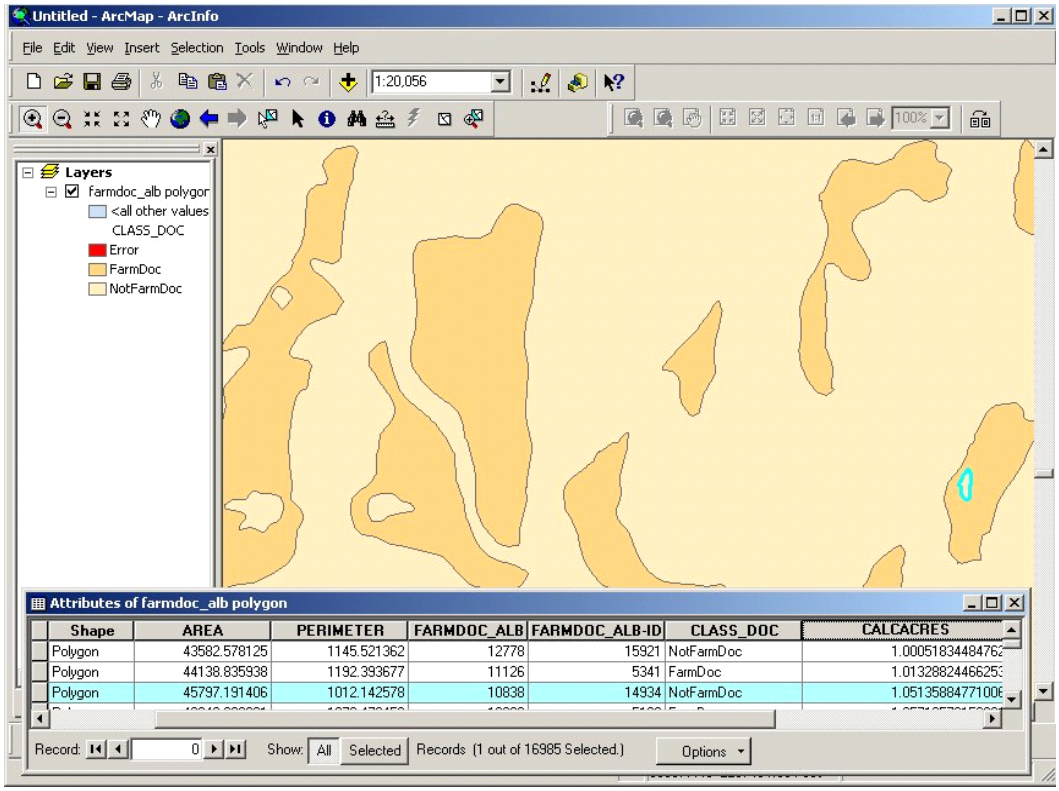
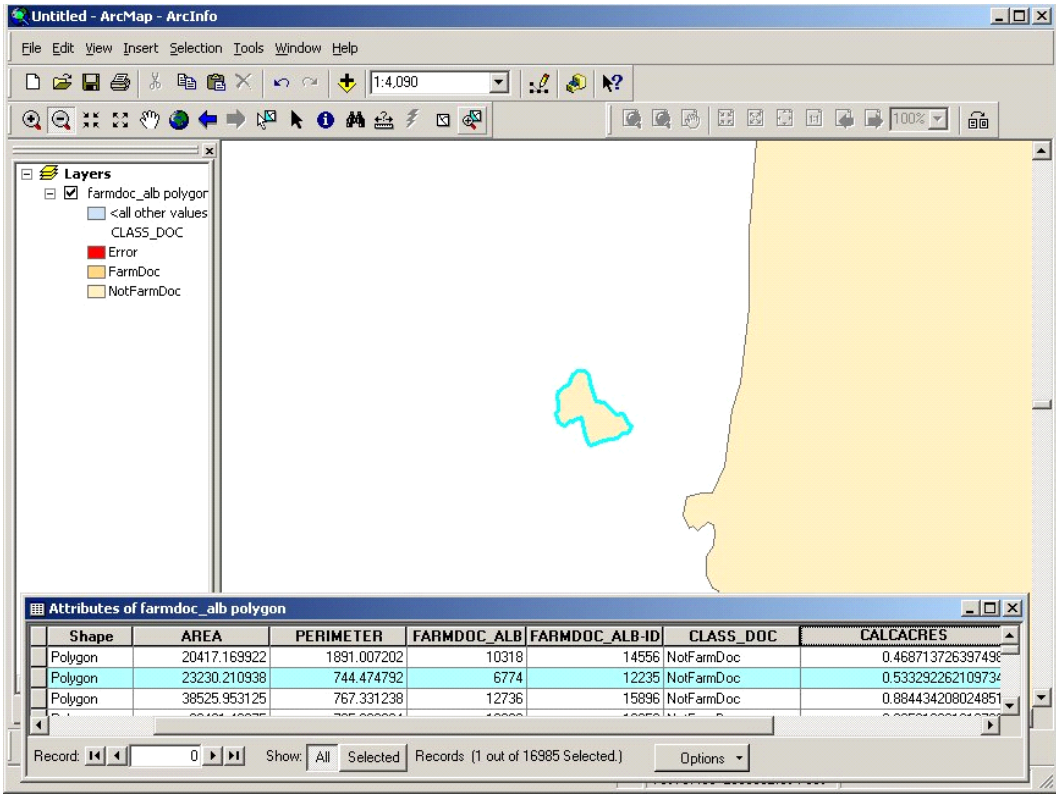


Figure 6. Example of "Not Farmland" polygon surrounded by a "Not Mapped" area



All of the above categories of sub-two acre polygons were removed by a reclassification process and then by dissolving them into neighboring polygons[18]. The code for this procedure is available in Appendix I.

As mentioned above, the process of compositing the data sets together resulted in the creation of undesirable artifacts. In this case, the term artifact is meant to describe those resulting polygons that are both small (generally smaller than 10 acres) and that originate from a general agreement and only slight spatial disagreement between the composite data.

The removal of these polygons was a nontrivial operation requiring the programming of many lines of code to identify those arcs from specified polygons that had a certain arrangement of polygons around them. For example, all sides from any FRAP polygon smaller than 10 acres were analyzed for adjacency to both a null classed polygon and an FMMP Farmland polygon. Code written to support this part of the process is available for review in Appendix I. This routine relied on a RELATE between the polygon arcs and adjacent polygons, both to the "left" and "right". There were mainly three types of artifact removed, mostly under 10 acres, but an additional approximate dozen were removed that were larger than 10 acres and as large as 41

[18] Simple deletion of polygons from ESRI Arc coverages will result in polygons with a null class field, which will get reclassed in any future union or other intersect operation.

acres. This last type of larger artifact was only removed where the depiction of farmland was unsupported by available aerial photography. The artifact types include:

(1) Stair-step shaped edges of FRAP polygons that were mostly overlaid by FMMP or LUSD data (Figure 7),

(2) Thin strips of FRAP that filled a sliver-type of gap between FMMP and LUSD (often at county boundaries; see Figure 8), and

(3) Irregular FRAP polygons that extend beyond the artificial boundaries of the FMMP and LUSD data, which may actually represent farmland, but which appear to be incongruous with the other data sources (see Figure 9).

Figure 7. Example of stair-step shaped edges

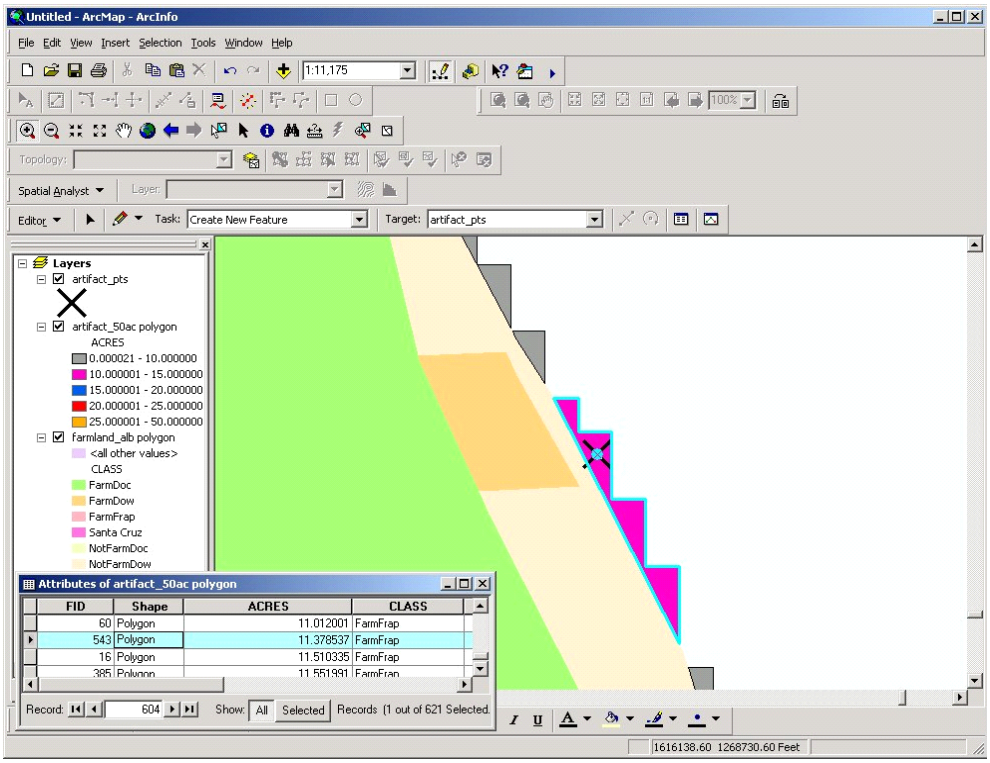


Figure 8. Example of thin strips of FRAP

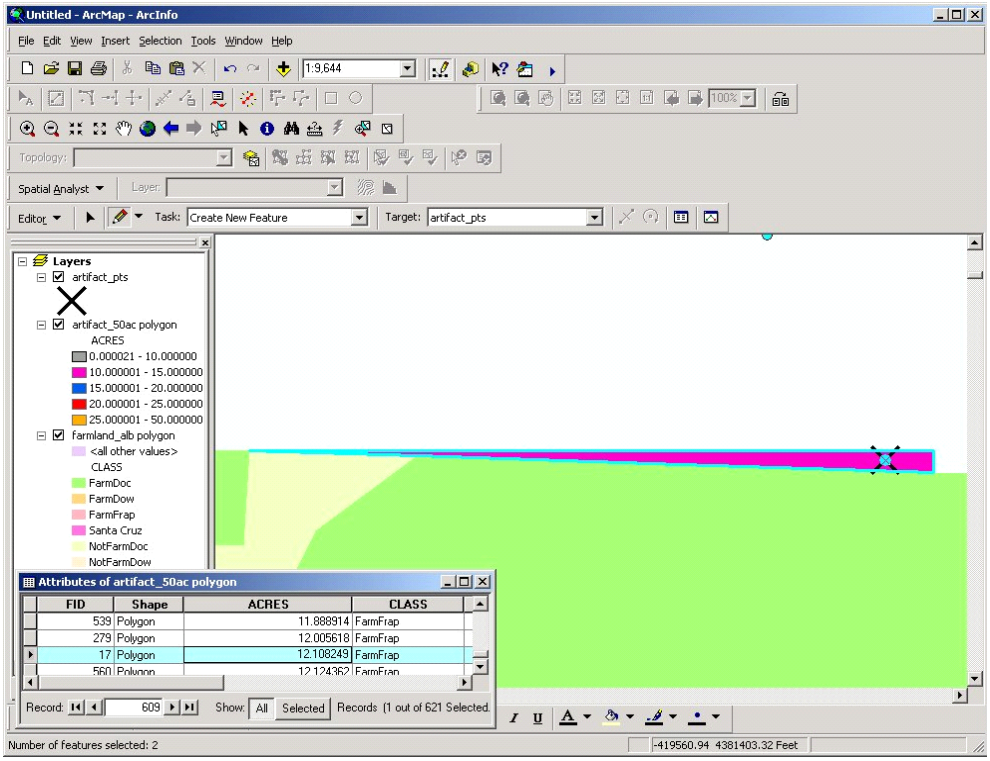
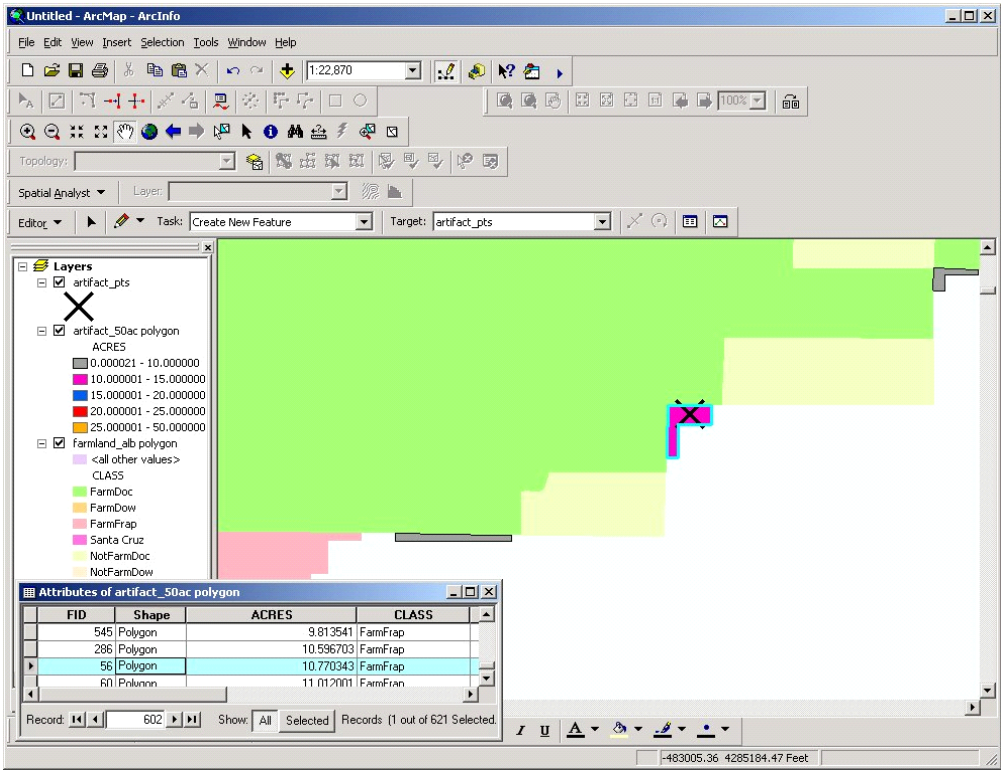


Figure 9. Example of irregular FRAP polygons



4.4 Characteristics of the Final Master California Farmland Data Base

The composite farmland data set represents a best effort at developing a single, statewide source of current farmland polygons for the State of California. Because no single source was available at the time of this compilation, the composite was generated. Decisions regarding which source should be authoritative in any area of overlap were required. The overlay hierarchy was established so that FMMP Farmland and Not Farmland always was selected preferentially to any other source, and LUSD Farmland and Not Farmland trumped was selected preferentially to any source other than FMMP. Finally, all farmland polygons in Santa Cruz County were replaced with the 2000-2002 Santa Cruz County Farm Parcels obtained from the county Agricultural Commissioner.

The sub-two acre polygons were removed in an attempt to drop most polygons from the FMMP, LUSD, and FRAP data that were below the reported minimum mapping units for those sources, but not so large that their absence (if actually valid) would greatly impact the overall validity of the final polygon data set.

Furthermore, due to the overlay and merge procedures required to composite the multiple data sets together, inevitably spatial artifacts were created where there was a mismatch in the depiction of farmland boundaries between competing data sets. The bulk of these artifacts were removed by developing a program to identify them and then delete them. However, some artifacts do remain as the removal procedures were designed to remove the largest amount of artifacts while deleting the smallest amount of valid polygons, which may happen to share characteristics with the artifacts.

Finally, a small set of polygons was individually marked for deletion after reviewing aerial photographs to determine that they were indeed probable artifacts and therefore candidates for deletion. Other members of the set were retained if the aerial photographs showed in any way that they were likely to be correct in their depiction of farmland.

4.5 Summary and Conclusions

Based on the unioning of multiple farmland data sources, and using GIS and the Arcinfo programming environment, a single comprehensive master farmland data base was compiled for an enhanced state-wide database of California farmland information. Due to the differences between data sources and software, including the ways in which ESRI shape files handle internal polygons and the way that ESRI's Arcinfo coverage handles them, special processing and discovery was required to ensure that unintended artifacts resulting from the data set union were removed.

A desirable future addition to this composite master data set would be the inclusion of source date information for each farm polygon. Due to the use of small polygon removal algorithms and the dissolution of internal boundaries within data sets between polygons with like-class categorizations (e.g., where Class = "DOC Not Farmland" for adjacent polygons within the DOC

or FMMP data layer), source date information might be best researched and appended to polygons in the original source data sets. Such information would be useful to users for many reasons, one in particular being the knowledge it conveys of whether the polygon distribution is likely to represent current farmed land use patterns.

Users of this composite data set should review the above discussion carefully, in addition to the metadata for the input data sources in Appendices D, E, and G.

5 CALIFORNIA PUBLIC SCHOOL LOCATIONS AND DEMOGRAPHIC DATA

An important component of the assessment is accurate information on the locations of the modeled receptor schools. This study acquired, reviewed, and updated location coordinates using several steps as explained below. The assessment also required retrieving and processing demographic characteristics of student populations by year to match the exposure averaging periods.

5.1 Universe of Schools

The first step in the process was to establish a single, definitive school inventory. Two data sources for schools (school year 1999-2000) were identified: the California Department of Education (CDE) school data base and the US Department of Education Common Core of Data (CCD). Quality assurance testing of the CCD data set revealed that, in addition to the known exclusion of Preschool Only Programs, the data set may have omitted as many as 45% of schools known to be operating at the time. Additionally, the CDE data set includes records that represent school districts and not schools. Since neither set alone contained either the full set of open schools in SY99-00, or provide a single source for reliable coordinate information, records from each were combined in an effort to create one comprehensive schools data set.

In selecting which schools to include in the inventory and which data set to use for overlapping records, the quality and presence of coordinate location information was considered, in addition to the availability of corresponding demographic information from the California Department of Schools. Some of the important characteristics of each data base are as follows.

CDE database:

- 10711 school records (where STATUS = pending or open, but not closed)
- Some coordinates provided only have 2 or 3 decimal places (indicating approximately an 800 meter error on the ground)
- Presumably includes programs that offer only preschool education
- Can include multiple records for the same school (or unique 7-digit school ID)
- Includes 1177 “district records” (see http://dq.cde.ca.gov/DataQuest/downloads/sifenr.asp for field definition)

CCD database:

- 8278 school records for California
- Excludes preschool only programs
- Approximately 85% of the records indicate that they were address-matched to the street address, as opposed to a zip code center point or other method, by the Agency for Toxic Substances and Disease Registry (ATSDR) using Geographic Data Technology
- 32 of the records do not contain latitude/longitude information
- Location coordinates do not appear to be truncated as with the CDE data

The unique school identifiers (i.e., the last seven digits of the CDS_CODE field in the CDE database and the 7-digit ST_ID field in the CCD) were compared to identify overlaps and discrepancies. Nearly 8,000 schools were present in both data sets. Since the CCD data set offers better location information, those records were used preferentially. The final inventory is composed of schools for which both location information and demographic information were available. Of those 8158 records came from the CCD data set[19] and 334 records from the CDE data set[20].

Once the school universe was assembled, it was checked to verify that priority schools were included. A priority school list, reported to be within 1.5 miles of at least 100 pounds of MeBr applications in 1998 (Ross and Walker 2000), was compared to the school universe. One school (Children's Center) was found to be on the priority list and not in the school inventory. This school was added to the inventory, although a school ID is not available and it appears to be primarily a limited-hours after-school program.

Following the development of the receptor schools database in 2004 (see Appendix H for a detailed description), an initial assessment of potential exposures was conducted for several sets of intermediate- and long-term exposure scenarios. The schools which were initially modeled with exceedances of target exposures were identified and further reviewed using aerial photography. Generally, it was felt modeled school locations should be confirmed to be accurate to within about 300 m of the schools identified through photos. Those with locations that had not been validated to this level were updated to provide locations inside the school's land parcel as seen in the photos. Subsequently, these updated locations were used in the final exposure evaluation. The file *schools.location.verifylist.xls* contains the set of schools reviewed, along with estimates of the accuracy of the modeled location, and revised locations for schools which did not meet the target accuracy.

5.2 Assessment of Original Location Information and Coordinate Acquisition

In this step, the coordinate information for each record in the school universe was assessed. Based on the fact that the CCD coordinates were of higher precision (more decimal places after

[19] Records omitted from the CCD data set are those that had no corresponding demographic information for the SY99-00

[20] Records were omitted from the CDE because: 1) demographic data were not available, 2) STATUS field was not "OPEN", 3) they had duplicate IDs, 4) school name and school ID were blank , 5) overlapped CCD data, or 6) were for juvenile detention centers, adult education programs, or district offices.

the zero) and that the records included documentation of an address-matching process which differentiated between records geocoded to street and to zip code, these coordinates were deemed more reliable than the truncated and undocumented coordinates provided with the CDE data set.

As a first cut, CCD coordinates were retained for all records geocoded to street addresses. The remaining set of 1558 records was sent to a commercial service for address matching to obtain location coordinates. These records included schools for which CCD coordinates were not available, schools for which the CCD had coordinates geocoded to zip centroids, and school records that originated in CDE.

At this stage, all coordinates were either from the CCD (count=~6900), or commercially geocoded (count=~1100). However, some of the commercially-obtained coordinates had still been geocoded to a zip code centroid (count=153), and CDE coordinates were substituted for these schools, except for those identified as a priority.

The identification of "priority schools" was performed to identify a set which may be more likely to be exposed to MeBr applications, and/or more likely to be influential in a disparity assessment. Initially, these priority schools included

- Those listed in Ross and Walker (2000) (count=410),
- Schools located in the same Census County Division as those listed in Ross and Walker (2000) (count= 3961), and
- High population schools (count=860; highest ten percent school population in the school universe)

Geographic Names Information System (GNIS) school location coordinates were used to replace as many of the coordinates for these high priority schools as possible. Approximately 250 priority schools were found in the GNIS. Because GNIS coordinates were developed from USGS 1:24,000 scale base maps, these coordinates were deemed more reliable than coordinates from these other sources which rely primarily on street address-matching. GPS-WAAS coordinates were collected for 5 priority schools, and an enhanced online address-matching service (Maporama) as well as Terrafly was utilized to obtain coordinates for a several additional schools.

For about 140 schools which initial modeling suggested may have been associated with relatively high concentrations, an additional location data verification step was performed. The previously derived school locations were displayed in digital aerial photos and evaluated. Revised coordinates derived from the imagery were created for 29 schools. A review of locations after the final round of exposure assessments revealed that three schools' locations were not within the target accuracy range of 300m, and were therefore excluded from the tabulation of potentially affected populations and disparity calculations.

5.3 Demographic and Enrollment Data

In order to estimate the number and characteristics of students potentially exposed at each school, demographic data was obtained. The CDE database (CDE 1995-2002) provides race and ethnic makeup information categorized[21] as

- African American not Hispanic,
- American Indian,
- Asian,
- Filipino,
- Pacific Islander,
- Hispanic or Latino,
- White not Hispanic, and
- Multiple.

The Hispanic or Latino category is the one of most interest to this study. The number or Hispanic and non-Hispanic students enrolled in each school was taken from the Demographics tables (Demographics 95-96 to Demographics 01-02) of the CDE data base. For most schools for school years 95-96 to 97-98, the number of Hispanic students was taken directly from the HISPANIC data field of the corresponding year's Demographics table, and the non-Hispanic enrollment was calculated by subtracting the number of Hispanic students enrolled from the total reported students (TOT_STDNTS).

Starting with the 1998-99 school year, however, the category "Multiple or No Response" was added by the CDE. To estimate the Hispanic enrollments for those school years, the percentage of Hispanic students was calculated from the sub-set of students whose race was reported for that year. That percentage was then applied to the total number of students at the school that year to estimate the total number of Hispanic students. Thus,

$$Hispanic\ enrollment_{year\ i} = [HISPANIC_{year\ i} / SUM(races)_{year\ i}] \times TOT\_STDNTS_{year\ i}$$

where

$SUM(races)_{year\ i}$ is the sum of the number of students reported in all racial/ethnic groups for that year at the school.

Additionally, at eight schools in at least one school year, there is no racial/ethnic data reported (i.e., all demographic fields are zero) but the TOT_STDNTS field is non-zero.

At these schools the percentage of Hispanic students was calculated for each of any other years that the school had reported demographic information, the percentages were averaged, and this average was then applied to the total number of students for the year of missing data to estimate the number of Hispanic students in that year. That is,

[21] The set of categories is not consistent with the 1990 US Census data, nor with the OMB directive 15, which distinguish more clearly between race and ethnicity as separate categories.

$$Hispanic\ enrollment_{year\ i} = AVG\_FR\_HIS \times TOT\_STDNTS_{year\ i}$$

where

AVG_FR_HIS is the average of the fractions of Hispanic students for each available year at the school.

The number of non-Hispanic students for both of the above cases was calculated by subtracting the calculated Hispanic enrollment from the TOT_STDNTS for that year.

6 METEOROLOGY DATA BASE

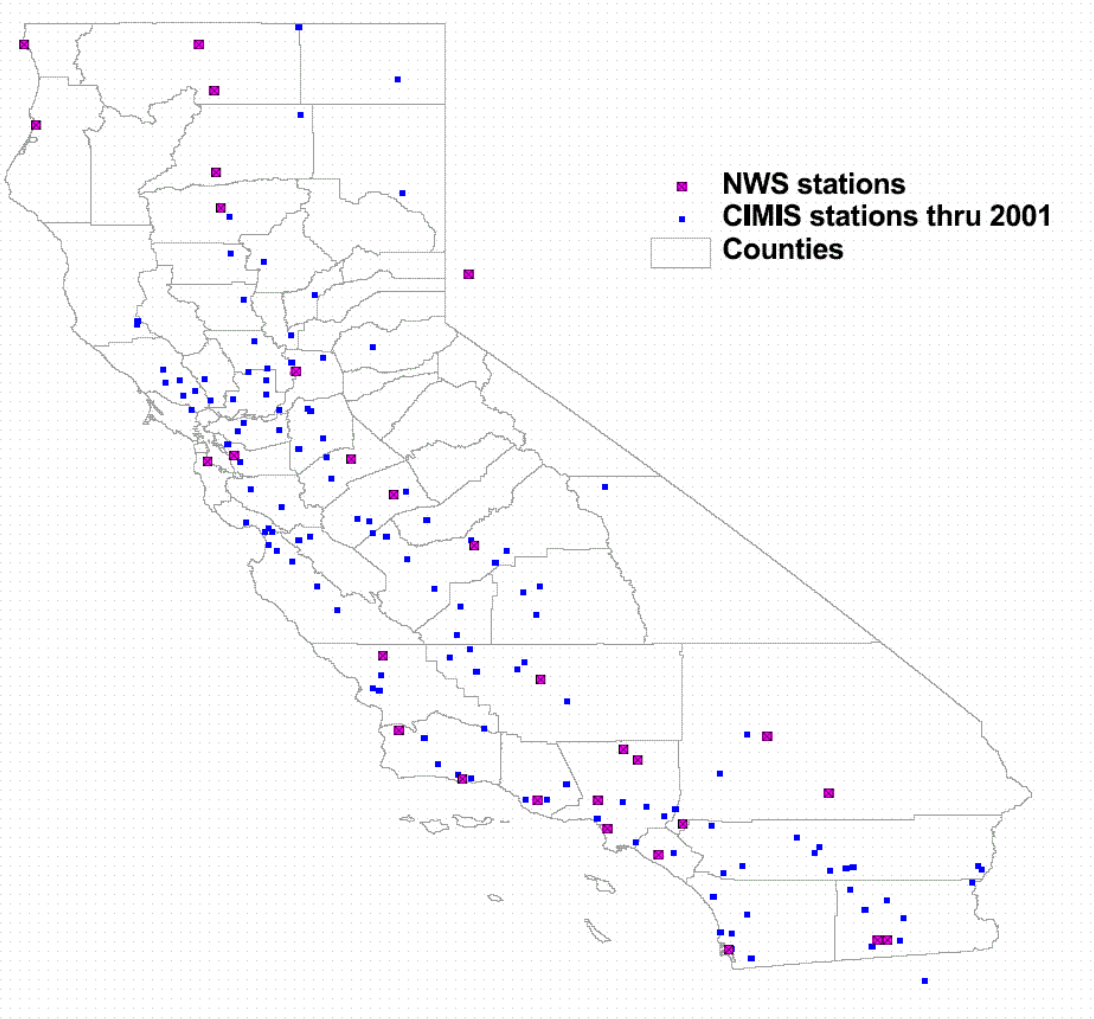
US EPA provides guidance for the use of meteorological data for air quality dispersion modeling in Appendix W to 40 CFR Part 51. The guidance specifies that the meteorological data selected should be spatially and temporally representative of the transport and dispersion conditions in the area of concern. To be representative the data should have the following characteristics:

- Collected at a location proximate to the area of study
- Collected at a location with similar terrain to the area of study
- Collected at a location with exposure to conditions over a wide area
- Collection at an elevation similar to the elevation of the emissions of the study
- Collected during the period of time of study
- Adequate quality control

In general, the use of site-specific meteorological data is preferred over data collected at a more distant permanent network station (e.g., National Weather Service or NWS) if a sufficient period of record is available.

California's Department of Water Resources has a statewide meteorological monitoring network, the California Irrigation Management Information System (CIMIS), designed primarily to assist farmers and others with the decisions about timing and magnitude of irrigation (http://www.cimis.water.ca.gov/). This extensive network of more than 100 meteorological stations covers all areas of California with a particular focus on the major agricultural regions. In this section we compare CIMIS meteorological data with National Weather Service (NWS) data for use in the MeBr regression model and describe the approach used in assigning a meteorological monitoring station to each California public school. Figure 10 shows the locations of the monitoring stations.

Figure 10. Map of CIMIS and NWS station locations



6.1 Evaluation of Criteria

6.1.1 Proximity of the monitoring location to the area of study

As noted above, the CIMIS network is designed to assist in making agricultural decisions, and thus the stations are located primarily in major agricultural regions of the state. In contrast, NWS data are primarily collected to support aviation and weather forecasting needs. The stations are generally located at airports and represent the surrounding area in the vicinity of the airport. Because the schools most likely to have MeBr concentrations of concern are located in the vicinity of agricultural operations, CIMIS stations are generally more proximate to the areas of most concern than are the NWS stations.

6.1.2 Similarity of terrain at the monitoring location to the area under consideration

In general both the CIMIS and NWS stations are located in areas where the influence of terrain on meteorological variables is minimal. For example, both station types avoid

topographic depressions, where the temperature is frequently higher during the day and lower at night than in surrounding areas. Thus, either type of station would be equally appropriate with respect to this criterion.

6.1.3 Exposure of monitoring location to meteorological conditions over a wide area

The placement of a CIMIS station is designed to accurately measure those meteorological conditions that influence evapotranspiration, including wind speed, temperature, net radiation, and humidity. Wind speed and temperature are variables used for the regression model. Nearby buildings or trees that may obstruct wind flows are avoided, as is the absence of green grass that can affect temperature, radiation, and humidity measurements. The ideal site for a CIMIS weather station is a 20-acre or larger pasture that is well maintained, with the sensor at least 100-yards from any wind obstruction.

NWS monitoring is designed to accurately measure conditions important for aircraft operations, including wind speed and direction, which can be used to evaluate wind shear; and temperature, which influences the length of aircraft takeoffs, landings, and fuel requirements, as well as the potential of icing conditions. Monitoring locations are similarly selected to be distant from obstructions to wind flows. However, NWS stations are typically in areas with a high level of paved ground cover, which can influence heat flows and, consequently the temperature. Thus, temperatures at NWS monitoring locations may differ from those in nearby agricultural areas for this reason.

Because CIMIS station locations have ground cover similar to the agricultural areas where MeBr emissions occur, CIMIS data would be preferred to NWS data with respect to this criterion.

6.1.4 Monitoring location elevation similar to that of the emissions of the study

Because evapotranspiration occurs primarily at or near ground level, CIMIS wind and temperature measurements are made near ground level at 2-m and 1.5-m.

In contrast, because evaluation of wind shear requires measurement of wind speeds and direction at elevations above the surface boundary layer, NWS monitoring typically occurs at 10-m above the surface, where the wind speed is generally higher.

Because agricultural emissions of MeBr to the atmosphere occur at the ground level, the near-ground-level meteorological measurements of the CIMIS stations would be preferred to those of NWS for this study.

6.1.5 Time period of measurements

Both the CIMIS and NWS observations are available for the entire seven-year period of interest, 1995-2001.

CIMIS calculates hourly averages from measurements taken once every minute throughout the day. Thus, the CIMIS data is an integrated average of conditions over prevailing over the entire hour.

Because NWS monitoring is designed to measure real-time conditions for weather warnings and forecasts, data are collected for a single one-minute period during each hour.

Because air transport and dispersion processes are continuous, the integrated average CIMIS data would be preferred to the "snapshot" NWS data for this project.

6.1.6 Quality assurance/control and maintenance

The level of quality control of the CIMIS data and the NWS data are similar. CIMIS data is reviewed by a quality control algorithm daily to compare measurements to historical means and standard deviations for each station, as well as theoretical limits. Additional quality control procedures are described on the CIMIS web site. On-site inspection and maintenance of sensors occurs every 3 to 6 weeks, depending on the season. Thus, either type of station would be equally appropriate with respect to this criterion.

6.1.7 Summary of Findings

Table 2 summarizes how well the CIMIS and NWS meteorological data meet the criteria listed above. As noted above CIMIS data would be preferred to NWS data with respect to the criteria of proximity to the study area, elevation of the measurement, and time period of the measurement.

Based on these findings we preferentially selected CIMIS data for the regression model application for locations that were sufficiently proximate to a CIMIS monitoring station. For other locations used data from a nearby NWS monitoring station if available.

Table 2. Comparison of CIMIS and NWS Data for Methyl Bromide Regression Model

	CIMIS			NWS		
Criterion	Ideal	Suitable	Minimal	Ideal	Suitable	Minimal
Proximity	X	X	X		X	X
Terrain	X	X	X	X	X	X
Exposure	X	X	X	X	X	X
Elevation	X	X	X			X
Time Period	X	X	X		X	X
QC/QA and Maintenance	X	X	X	X	X	X

6.2 Assigning Meteorological Stations to Schools

6.2.1 Criteria

For application of the regression model, each school needed to be paired with a representative meteorological station. Of particular importance to the analysis were the meteorological variables of wind speed and direction, temperature, and solar radiation (day/night indicator). The meteorological data needed to be representative of both the conditions at the

school as well as the potential emission source locations (i.e., farm fields). While it would be ideal to formulate an entirely objective procedure for use in evaluating representativeness, in general, no accepted analytical or statistical technique is available to determine representativeness of meteorological data (EPA, 2000).

The underlying assumption in our assessment is that meteorological conditions are constant within the space comprising the school and the corresponding emission source location, and during the time it takes for the emissions to travel from the source location to the school. Key factors which may cause spatial variations in meteorological conditions include:

- Distance between source and receptor: in general, the distance between the meteorological monitoring station and the school is the most important factor influencing the ability of the station measurements to be representative of the school location. The measurements are generally less representative as the distance increases.

- Water body influence: either both in areas strongly affected by on-shore air/sea boundary conditions, or both outside of such areas

- Height of the measurement: because atmospheric conditions can vary substantially with height, measurements from a station at a different elevation than a school may not be representative even if in close proximity. For this analysis meteorological stations which were located near ground-level were preferred, because the emission sources (i.e., soil fumigation activities) were located at ground-level.

- Terrain influences, which primarily effect wind speed and direction. These may drastically alter air flow patterns, even within a short distance between a meteorological station and a school, making the station measurements unrepresentative of the school location

- Similar surface characteristics as the school locations: similar surface roughness and land-use.

6.2.2 Decision Rules

Figure 11 presents the decision rules for pairing meteorological stations to each school in the regression analysis. In cases where professional judgment is specified the following factors were considered in making the assignment.

- Surface terrain features in the vicinity of the school and candidate meteorological stations
  - avoid pairing school with a meteorological station in a different valley;
  - minimize differences in valley axis orientation (e.g., up-valley station versus down-valley station).
  - avoid differences in water-body influences;
- Surface roughness features in the vicinity of the school and candidate meteorological stations: minimize differences in roughness values (e.g., farmland versus forested or built-up area);

- Proximity of the school and meteorological stations to urban areas: minimize differences in urban heat island influences

A preliminary review of the available CIMIS data showed nine locations within the state that had MeBr usage during the 1995-2001 time period, but are farther than 15 miles from any CIMIS station. Schools in the vicinity of those applications were paired with NWS meteorological data Altogether data from 117 CIMIS stations and 24 NWS stations were included in the meteorology data base. The names and locations of the 24 NWS sites are identified in Table 3.

Table 3. NWS Sites used for Regression Model Application

WBAN	Station	County	Latitude (deg N)	Longitude (deg W)
24286	Crescent City	Del Norte	41.78	124.23
24215	Mount Shasta	Siskoyou	41.32	122.32
24259	Montague	Siskoyou	41.78	122.47
24283	Eureka/Arcata	Humboldt	40.98	124.10
23258	Modesto	Stanislaus	37.63	120.95
23257	Merced	Merced	37.28	120.52
03159	Lancaster	Los Angeles	34.73	118.22
93184	Santa Ana	Orange	33.68	117.87
93121	Twenty-Nine Palms	San Bernardino	34.30	116.17
23199	El Centro	Imperial	32.82	115.68
03144	Imperial	Imperial	32.83	115.58
94262	Shelter Cove	Humboldt	40.02	124.07
23234	San Francisco	San Mateo	37.62	122.4
23136	Camarillo	Ventura	34.22	119.08
23273	Santa Maria	Santa Barbara	34.92	120.47
23130	Van Nuys	Los Angeles	34.22	118.48
23152	Burbank	Los Angeles	34.2	118.37
93209	Paso Robles	San Luis Obispo	35.67	120.63
93228	Hayward	Alameda	37.67	122.12
3179	Chino	San Bernadino	33.98	117.63
23182	Palmdale	Los Angeles	34.63	118.08
23155	Bakersfield	Kern	35.43	119.07
24257	Redding	Shasta	40.52	122.32
23185	Reno, NV	Washoe	39.48	119.77

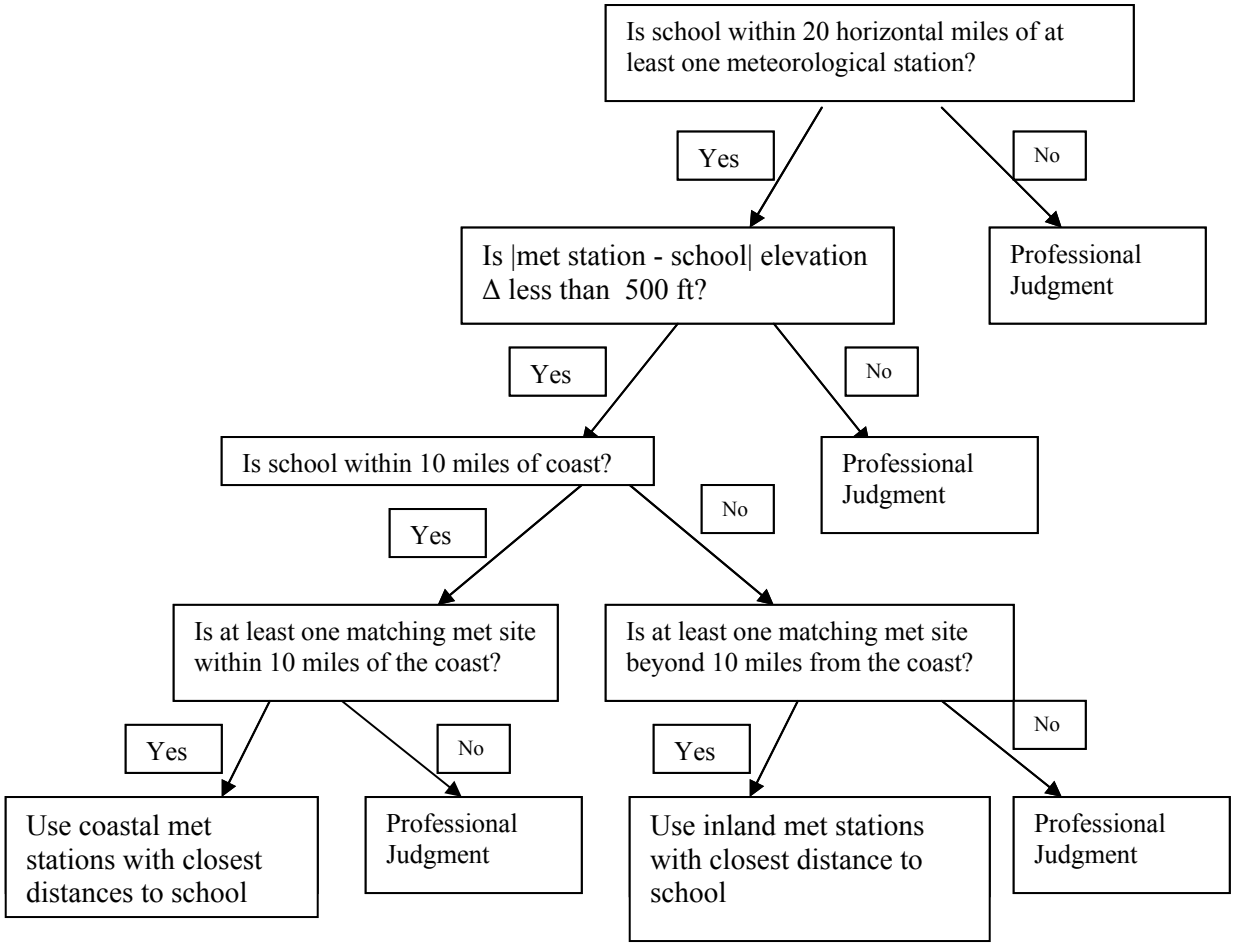
6.2.3 Primary and Secondary Pairing

For each California school except three, identification of both a primary and secondary meteorological station was assigned. The primary station was used in determining the daytime and nighttime wind and temperature values for input to the regression model, if at least 75% of the hours for each relevant time period have valid observations. Otherwise, the secondary meteorological station was used as input to the regression model.

Selection of primary and secondary meteorological stations for use in the regression model was made as follows. Each of 8,493 California schools was matched with the 140 meteorological

stations[22] using the objective criteria illustrated in Figure 11. For each school that achieved multiple objective matches, the two closest meteorological stations were chosen as the primary and secondary matches. Horizontal distances between objectively matched schools and meteorological stations ranged from about 0.1 to 20 miles. Of the 5,599 schools that achieved an objective primary match, 1,094 did not achieve an objective secondary match, and had a secondary match determined subjectively.

Figure 11. Decision assignment diagram for pairing of meteorological stations and schools



The remaining list of schools was first reduced by filtering for historical MeBr usage – any school whose location was greater than ten miles from the centroid location of any MeBr usage in the seven years between 1995 and 2001 This radius was chosen to extend beyond the boundary used in the regression modeling (8 miles), for which the concentration would be set to zero.[23] There were 1,386 schools that had no historical MeBr usage within 10 miles and were

[22] CIMIS station 005, Kern County, San Joaquin Valley was removed from consideration due to lack of adequate wind direction measurements

[23] The regression model developed for this study does not include a background concentration, i.e., the intercept was constrained to be zero. This was done because an unconstrained intercept value would represent the unique proximity of the monitor locations to background sources, and thus it is very likely to be spatially variable. If we included an intercept in our modeling it would represent the typical background concentration in the region where

excluded. The remaining 1,508 schools had both a primary and secondary meteorological site chosen subjectively.

Stations requiring subjective selection were paired primarily by proximity, but respecting each of the criteria listed above. Generally, stations were allocated based on sharing the same meteorological characteristics, determined by proximity, sharing the same inland or coastal climate, being located in the same region of the same valley (if applicable), having similar altitude, having similar exposure, and having similar terrain. Each of the schools requiring subjective judgment was assigned at least a single match that met all these requirements. All matches were assigned a quality score from 1 to 4 based on how well they met these criteria, with 4 being an objective match.

The first full set of school-meteorological station pairings, including objective and subjective primary and secondary matches, was passed through a quality check to verify that there was sufficient data from each meteorological station pair to properly complete the regression modeling. The feedback from this indicated that a relatively small number of meteorological stations (20) were contributing the majority of the gaps in the paired data (about 76 %). In cases where these twenty stations appeared as a second choice, they were substituted with another station that adequately represented the same region. None of the primary sites were changed. This sufficiently resolved the majority of meteorological data gaps.

6.3 Solar Radiation and Day/Night Indicator

For each meteorological station an hourly day/night indicator is calculated for use in the regression model. The day/night indicator is based on the location of the meteorological station, time of day, and day of year. A daytime flag is specified for a particular hour if the solar elevation angle is greater than 10 degrees above the horizon. The algorithm for calculating the solar elevation angle is the same as that used in EPA's Meteorological Processor for Regulatory Models (MPRM) (EPA, 1996).

7 DISTANCE AND ANGLE TABLES

As noted above, the final data base developed for this study was a set of tables generated using GIS tools and spatial analysis for the purpose of estimating the approximate distance of an application from the school receptor location and the angle between the application and the school relative to the wind direction. Interim GIS data sets used to create these tables included the Meridian Township Range Section (MTRS) grid for California, a customized ¼ mile by ¼ mile grid, the farmland data base, the school data base, the meteorology data base, and MeBr usage data reported at the MTRS level.

---

the monitoring data on which the model was based was collected, but could be a very poor estimate of the background concentration in other regions

7.1 Finer Scale Model Grid

In choosing a finer scale model grid cell size, a number of factors were considered: (1) the average farm field size, (2) the fate of applied MeBr, and (3) computer resource requirements. A number of grid cell sizes were benchmarked, in terms of performance, including side lengths of 1/10 mile, 1/8 mile, ¼ mile, and ½ mile. The ¼ mile grid cell size (40 acres, or 1/16 of a square mile) was ultimately selected for being the smallest feasible cell size to work with. While there are about 160,000 MTRS sections that cover California, there are more than 2.5 million ¼ mile grid cells required to cover the same area. Sixteen ¼ mile grid cells fit into one MTRS section. GIS files for gird with 1/8 mile or 1/10 mile spacing (10 acre or 6.4 areas, respectively) exceeded the approximate 2.1 Gigabyte file size limitation for a 64 bit application.

7.2 "Pseudo-Farmland"

The development of the farmland data base is described in section 4. One limitation of the farmland data base is the small number of mismatches where the CDPR reports usage in an MTRS but no farmland was identified in the MTRS. These discrepancies are likely due to the fact that agricultural land use changes over time, while the data bases from which our farmland data set was derived represent a series of temporal snapshots collated together. Many of the more than 200 such instances were investigated, especially for cases where the MTRS was in close proximity to a school and/or the level of reported usage was high. In some cases, field-specific farm parcel data were obtained from the county agricultural commissioners to supplement the farmland data base for use in calibrating and applying the regression model. (See section 7.4.2 below). In other cases, it was determined that the usage had been reported to the wrong MTRS, and the usage data base was corrected.

For the remainder of MTRSs with reported usage but no identified farmland, the MTRSs are treated as "pseudo-farmland" in the reapportioning routine, with the usage distributed evenly across the entire approximately one square mile area, i.e. the entire MTRS is assumed to be farmland. While this approach may somewhat dilute the modeled concentration estimate of any single high level application of MeBr, it is preferable to excluding the usage from the model application for lack of information on its exact location.

7.3 Spatial Reapportionment of Usage Data

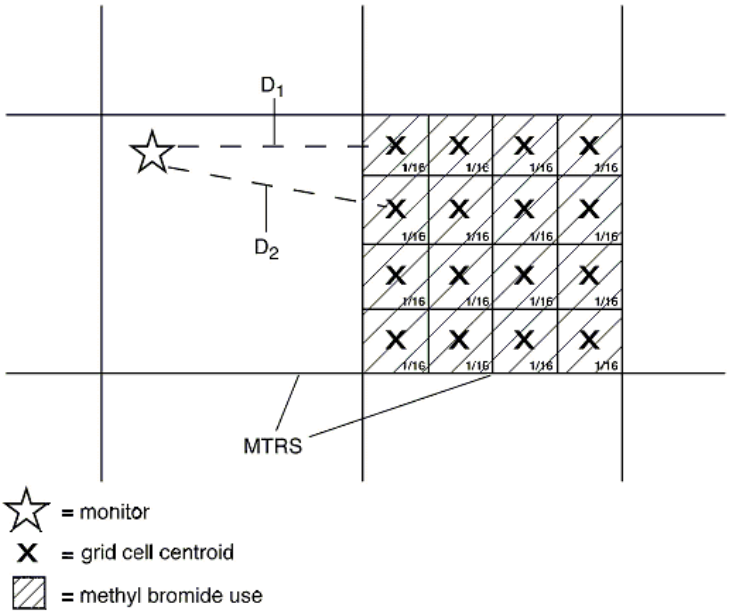
A substantial portion of the project activity was devoted to improving the geographic locations of the MeBr usage data. Since the PUR data include only the MTRS section ID, one approach (used by CDPR) is to assume an equal distribution over the entire section, represented as a point at the center of each section. However, MTRS areas are usually large areas of about 1 mile square (approximately 640 acres). Since the regression model process treats the use as occurring at the center point to represent an entire MTRS section, such an approach assigns all of the usage to points located about 1 mile apart in the majority of cases.[24]

[24] In some instances, MTRS sections are much smaller or irregular in size, especially when located near meridian boundaries. For these exceptions, the grid points would be located an unpredictable distance apart.

An approach that would be expected to improve the spatial resolution of usage estimates is to overlay a single MTRS section over a regular grid of multiple cells, and assign the usage across the associated cells. Instead of locating usage at a single point at the center of each section, this approach spreads it across a group of cell center points, which more closely approximates the distribution across the entire section. This decreases the error in use location resulting from assigning all usage to a single point within a section, and was one aspect of the approach adopted in this study. The grid cells used are approximately 40 acres in area, and a typical MTRS section would contain 16 cells. (See Figure 12). For this approach the number of points that represent source locations near a receptor can be large, i.e., about 16 times as many as for the MTRS-center-points approach. For the maximum extent of contributing sources examined here (8 miles radius, or approximately a 16x16 mile area of MTRS), the total number of potential source points is about 4000 for each receptor.

Figure 12. Illustration of reassigning methyl bromide usage data from approximately 1-mile-square MTRS's to ¼-mile-square grid cells in equal proportion as indicated by the fraction in the corner.

(Note that the actual MTRS sections are sometimes irregular, and that the grid cell edges do not align with the MTRS boundaries.)

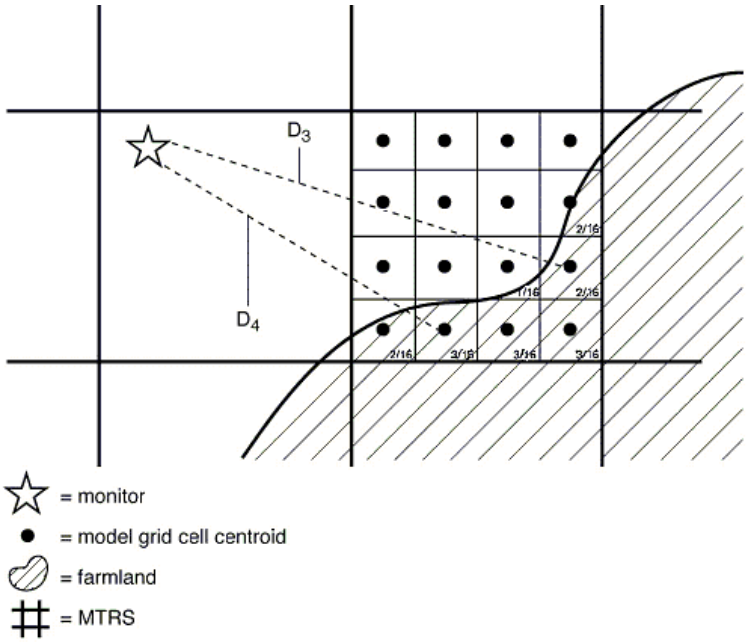


This project also goes further, to use additional data to apportion usage occurring within an MTRS to only the cells containing suitable land cover categories (i.e., farmland). This is similar to the approaches described in Rull and Ritz (2003) and Ward et al (2000), and builds on available California farmland data to use as a mechanism to weight the allocation of usage data to associated grid cells. This approach more closely approximates actual field-specific usage location because it confines the usage to cells containing farmland. The proportion of each MTRS section's total farmland within a grid cell is used to apportion that section's usage accordingly.

Geographic Information System (GIS) mapping software was used to assign the usage to square grid cells of size ¼ mile on each side. For each MTRS section, the total amount of farmland was summed across all of the square grid cells intersecting that MTRS, and the daily usage for each of those grid cells was assigned in proportion to the amount of farmland in that grid cell.[25] (See Figure 13.) Because many MTRS sections are not perfect 1-mile x 1-mile squares, the grid cells do not line up with the MTRS boundaries, so that a grid cell may be assigned usage from up to four MTRS sections. The small and irregular MTRS sections near the Mettler Fire Station (MES) monitoring site were able to be processed using this method. This allowed the inclusion of this site in this study's model calibration process. In a few cases outside the area near the monitoring sites, an MTRS section with some reported usage had no farmland identified in our farmland database; in those cases the usage is apportioned uniformly across all the grid cells intersecting that MTRS.

Figure 13. Illustration of reassigning methyl bromide usage data from approximately 1-mile-square MTRS's to ¼-mile-square grid cells, in proportion to the amount of farmland contained in each grid cell, indicated by the fraction in the corner.

(Note that the actual MTRS sections may be irregular, and that the grid cells do not align with the MTRS boundaries.)



For each grid cell and monitoring site, the GIS was used to calculate distance and angle from the center point of the grid cell to the monitoring site. The models can then use the distance between the actual monitoring site location and the center point of a small grid cell (with an area

[25] Geographic information about farmland in California was obtained from three sources: the California Department of Conservation, the California Department of Water Resources, and the California Department of Forestry and Fire Protection's Fire and Resource Assessment Program (FRAP).

of about 40 acres) in the prediction of monitored concentrations as a function of distance. The angle information is used in conjunction with wind direction data for some model formulations. This analysis still makes the approximation that all usage occurs at the center of the grid cell, neglecting the distribution over some fraction of the actual cell area, and the errors due to the fact that an individual fumigant application would be applied only to some of the farmland inside that grid cell.

As discussed in section 7.4.2 below in some cases, we were able to obtain detailed location information on the farmland parcels where the MeBr was applied. These field-specific application data were used to identify to which specific grid cells to assign the usage. For the records that could be matched to the detailed farmland parcel information, usage was assigned to each specific field receiving an application of MeBr on a particular day. For those records, usage for a field was apportioned proportionally to each grid cell based on the area of the field's farmland overlapping the grid cell.

But for most records outside the set for which we obtained specific farmland parcel data, usage for an MTRS was apportioned proportionally to each grid cell based on the area of the MTRS's non-specific farmland in the grid cell, as described above and in Figure 13.

7.4 Algorithm for Spatial Reapportionment the MTRS-based Usage Data to the Finer Grid

In order to reapportion the amount of MeBr usage from the MTRS to the finer scale grid, the classic proportional overlay procedure was applied with GIS tools. This procedure calculates the fraction of the area of the MTRS section that is co-located with (or "overlays") any portion of any given finer scale grid cell. However, in order to apportion usage only to those areas that are likely to have had MeBr applied to them, only those portions of the MTRS that are classified as "farmland" according to the farmland data base assembled for this study were considered. Although MeBr is not necessarily applied to all farmland, this approach ensures that areas with very low probability of MeBr application (i.e., those not is agricultural use) are excluded from the analysis.

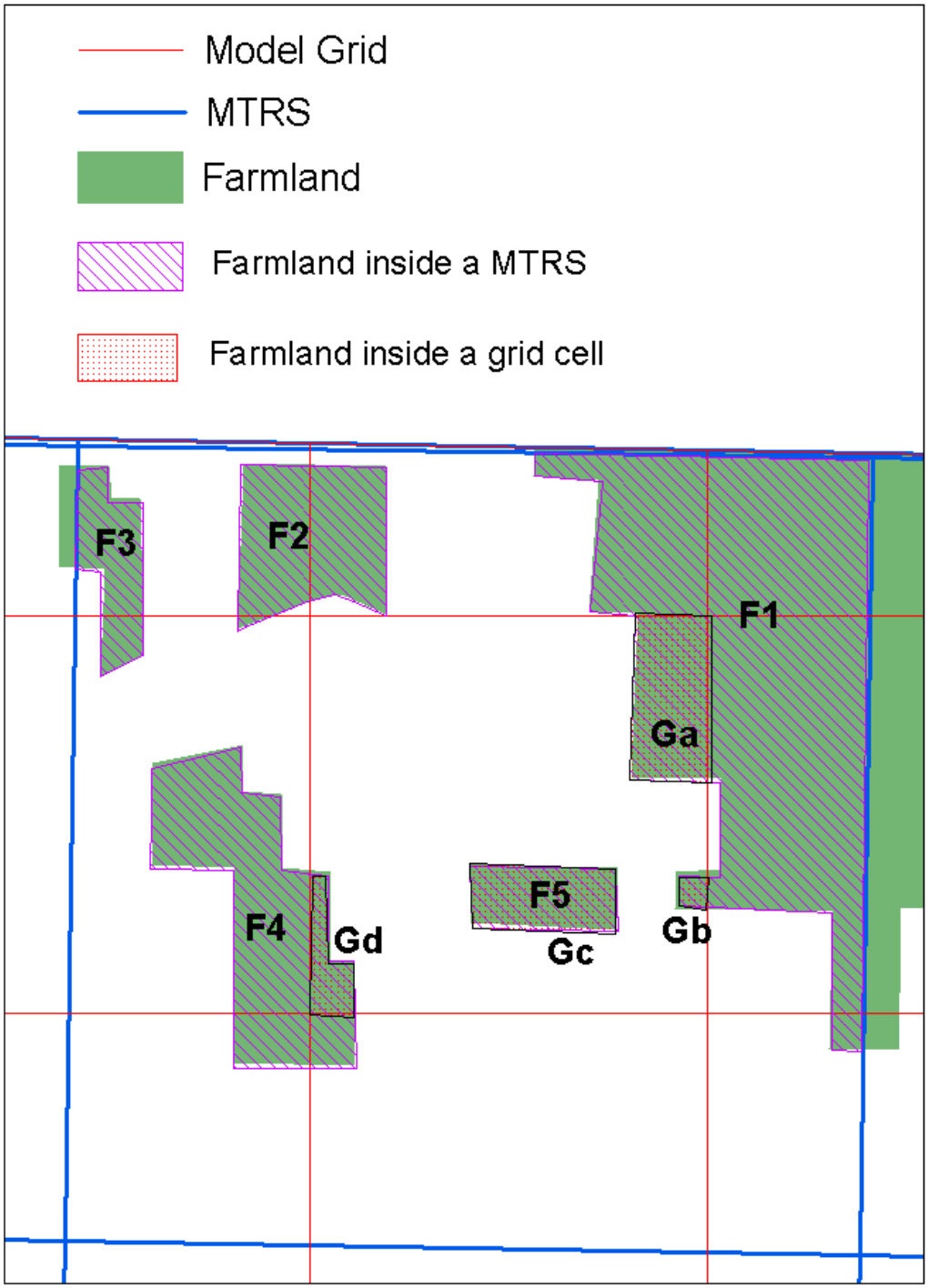
Figure 14 shows an example of how the overlay procedure works. Based on the annotated polygons in the figure, the fraction of the total MTRS usage that should be apportioned to the finer scale model grid cell is calculated with the following formula:

$$(Ga+Gb+Gc+Gd)/(F1+F2+F3+F4+F5)$$

where F1 through F5 represent the areas of the farmland polygons that intersect the MTRS and Ga through Gd represent the areas of farmland polygons that intersect the constituent grid cell.

Appendix J presents the script used to calculate the various pieces of this formula iteratively across the entire study area and Appendix K presents a sample of the output table produced. An important assumption underlying this approach is that the MeBr is applied at the same rate across the constituent farmland areas. This assumption may lead to an underestimate of peak exposure concentrations, and an overestimate of the spatial extent of MeBr exposure.

Figure 14. Example of overlay procedure to apportion MTRS usage to farmland within model grid cells.



7.4.1 Vector vs Raster Environment

Because of the large amount of data involved in this analysis, use of a raster GIS environment was considered. Raster data models provide faster iterative data processing, due to the uniform size of the minimum mapping unit. However, a vector environment was selected because (1) most of the data were already in vector format and conversion to raster with related quality assurance would be resource intensive, and (2) the accuracy requirements for area calculations would entail a small minimum mapping unit in a raster environment, which in turn would slow processing.

7.4.2 Field-Specific Data

In the process of calibrating the regression model (Cohen et al. 2011), it was found that single high usage applications could have a great influence in the exposure surface for a large area around the application location. Therefore, it was important that high usages reported in the vicinity of complaint schools, in particular, be mapped to actual fields as accurately as possible.

We also tried to find field-specific data if the generic farmland data did not identify any farmland or only a very small amout of farmland in an MTRS where MeBr usage was reported to occur (see discussion of "pseudo-farmland" above). These efforts were undertaken to try to minimize the use of "pseudo-farmland" in the re-allocation process.

Field-specific maps were collected for MeBr usage from Santa Cruz, Monterey, Ventura, and Kern Counties[26]. Several such maps had already been obtained and digitized for areas with large usages in the vicinity of locations of ambient monitoring data used to calibrate the regression model (Cohen et al. 2011). Due to the effort involved in requesting, finding and then digitizing field maps, additional maps only in high priority areas were requested. Generally, these were areas that were either

- within 3 miles of a complaint school, or
- reported greater than 4000 lbs of usage per MTRS, and had less than 2% farmland acreage.

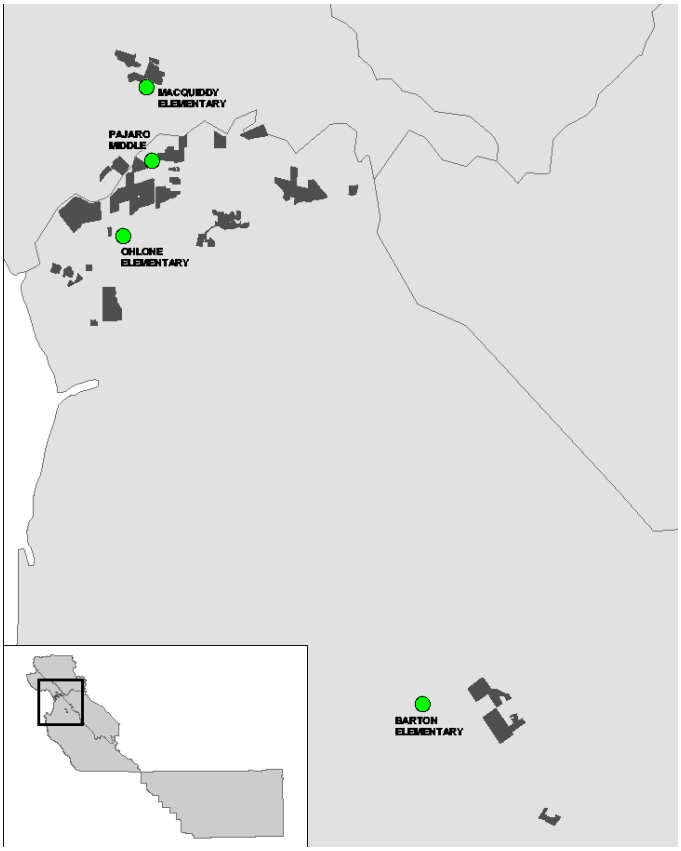
These requests were submitted to the counties in an effort to collect associated field maps for the usage reports. Although all maps that were received were digitized and included in the modeling application, maps were not provided for all fields, or all years of usage, in the high priority areas.

[26] While the PUR database contains geographic coding to the MTRS, it also contains references to specific field identifiers. These are not unique, in that they can be repeated for different growers (field 01, 02, etc). In combination with a grower ID, they can often be linked to specific computerized polygons digitized from paper maps. Several California counties are actively involved in developed geographic information system (GIS) data layers of fields receiving pesticide applications, including Monterey and Santa Cruz counties (Neal, 2001 - slide 17). Moreover, many of the paper copies of the pre-application Notice of Intent (NOI) and/or post-application Pesticide Use Report (PUR) records submitted to county agricultural commissioners by applicators contain field maps with identifying features, as well as the specific day(s) of application.

Santa Cruz supplied full documentation and field polygons, already in GIS format, for the years 2000 and 2001. The other counties responded with paper maps, in some cases hand-drawn, with few location references, and sometimes with erroneous MTRS identifiers. Appendix L lists the identification parameters for the field-specific usage data. Appendix M contains the contact information for the County Agricultural Commissioners that assisted us with this study.

Although the boundaries for these fields were not integrated into the master farmland dataset, they are available in GIS format separately. Figure 15 shows the first 40 of the fields that were digitized.

Figure 15. Example of overlay digitized farm fields.



In order to redistribute the MeBr usage data to the digitized farm fields, it was necessary to match the field boundaries to the usage report records. Because a single unique ID was not available to link these across time (field boundaries were available for areas where MeBr was applied for any year between 1995 and 2001) the following steps were taken to match the spatial records to the usage reports.

1. Farm fields were digitized using street intersections and other markings on the hardcopy map to register the map to Earth coordinates. Due to the limited amount of markings on the hardcopy maps, some fields were difficult to spatially locate with complete confidence.
2. The digitized farm fields were compared to polygons already in the master farmland dataset.

3. Application year(s) were noted for each digitized farm field. (The application year is available as part of the grower ID.) Notes were kept for each field regarding confidence levels in application year assignment.
4. A unique polygon ID was assigned to each digitized farm field.
5. The farm field polygon ID was entered into year-specific fields in the usage table. This step linked the farm field to each usage record for each applicable year.

Figure 16 displays an example of specific farmland parcels in Ventura County, with MTRS IDs and MeBr usage displayed by section. The maps associated with the usage reports also reflected only single years, so that the re-apportionment for these locations was conducted on a year-by-year basis with reported MTRS usage assigned to specific fields. In locations where field-specific information was not available or digitized (the vast bulk of the state) reapportionment was conducted for the entire period 1995-2001, with reported MTRS usage applied uniformly across farmland present in the MTRS.

Figure 16: Example field-specific farmland data in Ventura County

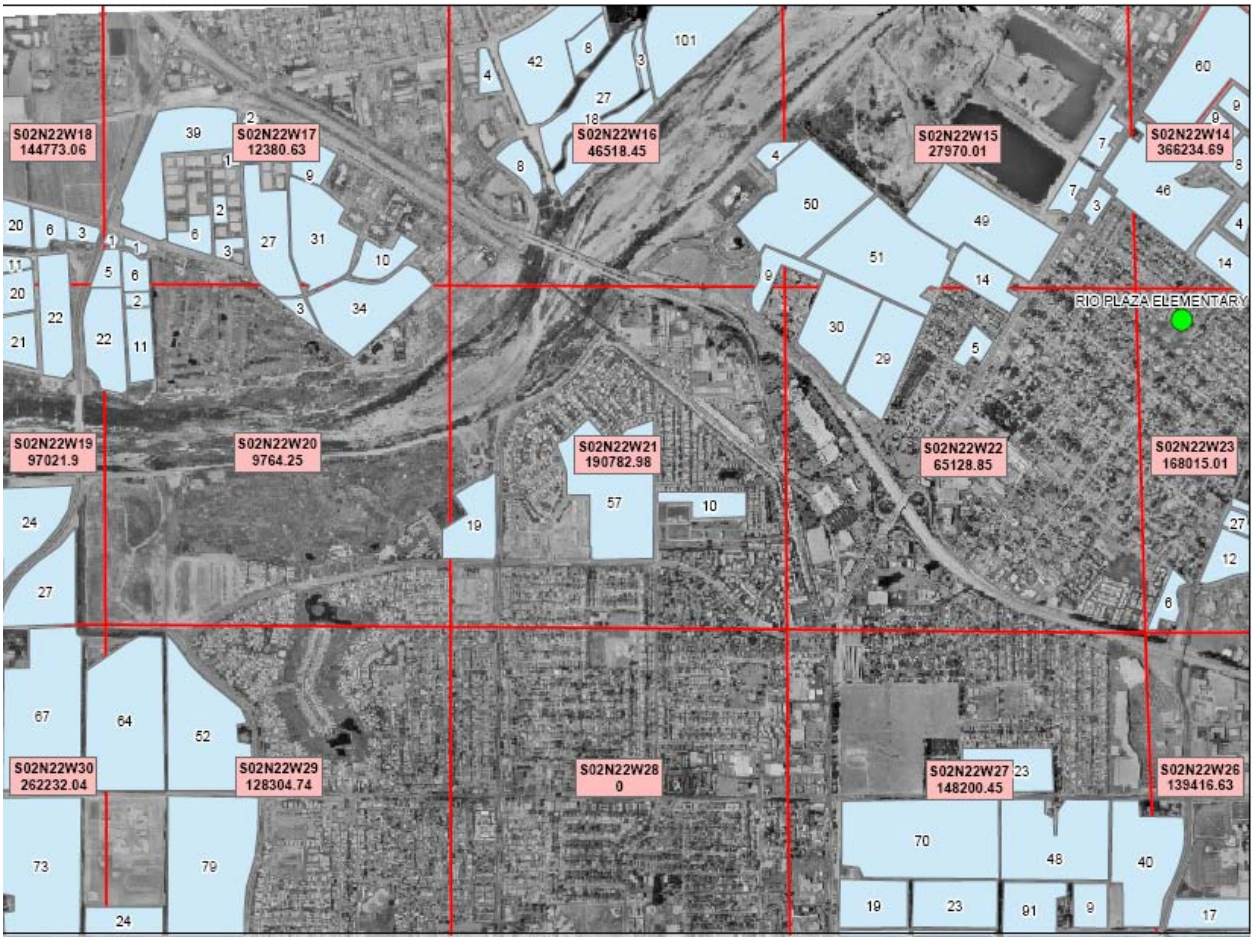


Table 4 and Figure 17 summarize the proportion of total MeBr usage within 3 miles of each complaint school that was linked to specific farm fields.

Table 4. Percent of Total Usage Within 3 Miles of Complaint Scools Linked to Field-specific Data

	Pajaro	MacQuiddy	Ohlone	Barton	Rio Plaza	Rio Mesa
1995 - 1997						
Total Usage	1,954,224	1,783,271	1,942,426	1,090,658	1,196,453	1,132,179
Field-specfic Usage	426,813	428,700	380,986	167,175	197,179	197,179
Percent Captured Usage	22%	24.%	20%	15%	16%	17%
1998 - 2001						
Total Usage	2,621,176	2,324,971	2,585,661	1,281,235	2,288,679	2,167,209
Field-specific Usage	1,423,229	1,379,973	1,090,353	115,590	773,057	773,732
Percent Captured Usage	54%	59%	42%	9%	34%	36%

7.4.3 Investigation of Usage Records with Blank or Erroneous MTRS Identifiers

As noted above, PANNA initially provided MeBr usage data for the entire state. A number of records lacked identifiable MTRS IDs. It is likely that many of these missing or inaccurate IDs were missing from the original usage reports submitted by permittees, since only one or two grower IDs in each county were listed in the records. Although the vast majority had zero or very limited associated usage, four records had over 10,000 lbs of usage and 35 had over 3000 lbs of usage. An effort was made to investigate these high usage records to identify the correct MTRS locations. The Department of Pesticide Regulation's CalPIP application was used to crosscheck the usage records for verification. Several records were updated with a valid MTRS, and the omission of the other records from the assessment will have a negligible effect on exposures (with the possibility of slight underestimates as a result of the missing values).

7.5 Quality Assurance on Area Calculations

Because millions of records were being operated on via proportional overlay, involving millions of polygons, an automated means of verifying the accuracy of the results was required. In Appendix N the ESRI ArcObjects ® code developed to aggregate and sum the results in stages, is presented. The aggregating is performed by year in the case of the field-specific calculations. Additionally, "pseudo-polygons" are counted and reported on during this quality assurance program. Appendix O presents a sample output from this program, as run for the pilot test that covered the 2000-2001 time period.

7.6 Development of the Distance and Angle Table

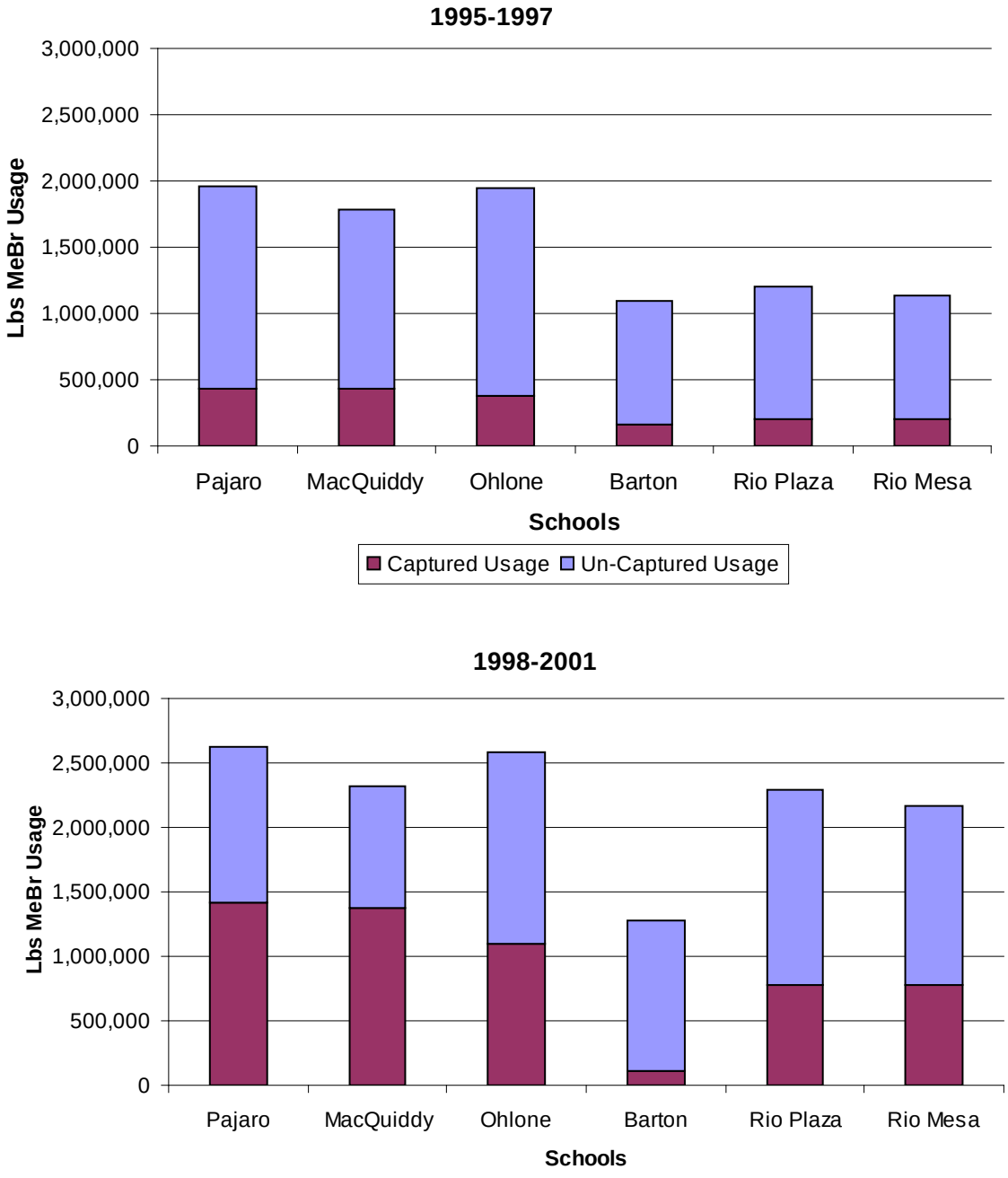
The distance and angle between each California school and every ¼ mile grid cell's center point within 8 miles[27] was calculated in order to support the application of the regression model developed for this study (Cohen et al. 2011) to estimate MeBr concentrations at California public schools.

GIS tools were used to support all steps in the creation of these tables. The MTRS grid layer, model grid layer, farmland, and school location layers were collected and registered to the same coordinate system. They were then spatially stacked in order to be able to calculate their relative

[27] Eight miles is the outermost boundary for MeBr applications used in the regression model (Cohen et al. 2011).

positions (distance in meters and angle in degrees between schools and the ¼ mile model grid center points). Appendix P presents the code used to produce the tables. A sample of a distance and angle table is presented in Appendix Q.

Figure 17. Percent of total usage within 3 miles of complaint schools linked to field-specific data



8 REFERENCES

CDE. 1995-2002. California Department of Education. Downloadable data files for California Basic Educational Data System (CBEDS) enrollment by ethnic group and school. http://www.cde.ca.gov/ds/sd/sd/filesethsch.asp

Cal-EPA/CDPR. 2000. *Pesticide Use Reporting; an Overview of California's Unique Full Reporting System*. May 2000. http://www.cdpr.ca.gov/docs/pur/purovrvw/ovr52000.pdf

Cal-EPA/CDPR., 2002. *Pesticide Use Report Data: User Guide & Documentation* Department of Pesticide Regulation, Information Systems Division. http://www.krisweb.com/biblio/cal_dpr_dpr_2002_pesticidecdguide.pdf

Cohen, JP, Rosenbaum, AS, and Carr, EL. 2011. *Model Development for Assessing California Methyl Bromide Ambient Concentrations*. Draft report prepared by ICF International for Loren Hall, EPA Office of Civil Rights, Washington DC.

EPA. 1996. *Meteorological Processor for Regulatory Models User's Guide*, Office of Air Quality Planning and Standards, EPA- 454/B-96-002, August 1996.

EPA. 2000. *Meteorological Monitoring Guidance for Regulatory Modeling Applications*, Office of Air Quality Planning and Standards, EPA- 454/R-99-005, February 2000.

Neal, R.H., 2001, *The Use of GIS in Reporting Pesticide Applications in California (Briefing)*, CDPR, http://www.cdpr.ca.gov/docs/county/pumpdvlp/presents/calgis

Nordstrom, J., 1996, Procedure: PLSNET Coverage Final Form, http://www.ccjdc.org/resources/WaterData/pesticid/plsnet/plsfinal.doc

Personal Communication, January 10, 2003, Susan Kegley, Pesticide Action Network, Provided: California 2001 Methyl Bromide Usage Data by Meridian Township Section Range.

Personal Communication, May 5, 2003, Susan Kegley, Pesticide Action Network, Provided: California 1995 Methyl Bromide Usage Data by Meridian Township Section Range.

Personal Communication, June 23, 2003, Stephan Orme, Pesticide Action Network, Provided: California 1996-2000 Methyl Bromide Usage Data by Meridian Township Section Range.

Personal Communication, October 1, 2003, Susan Johnson. Ventura County Assistant County Agricultural Commissioner

Personal Communication, October 2, 2003, Charles Kregel. Riverside County Pesticide Officer for Cochallea Valley.

Personal Communication, October 8, 2003, David Case. Tulare County Deputy County Agricultural Commissioner.

Personal Communication, October 30, 2003, Paul Niday, Vice President of Operations TriCal, Inc., Hollister, CA.

Personal Communication, October 3, 2005, Susan Kegley, Pesticide Action Network, Provided: California 2002 and 2003 Methyl Bromide Usage Data by Meridian Township Section Range,

Ross, Z and Walker, B. 2000. *An Ill Wind: Methyl Bromide Use Near California Schools 1998*. Environmental Working Group. http://www.ewg.org/files/anillwind_illwind.pdf

Rull, Rudolph P. and Ritz, Beate. 2003. "Historical Pesticide Exposure in California Using Pesticide Use Reports and Land-Use Surveys: An Assessment of Misclassification Error and Bias." *Environmental Health Perspectives*. Vol. 111, No. 13. October 2003. http://ehp.niehs.nih.gov/docs/2003/6118/abstract.html

Ward, Mary H, Nuckols, John R, et, al. 2000. "Identifying Populations Potentially Exposed to Agricultural Pesticides Using Remote Sensing and a Geographic Information System." *Environmental Health Perspectives*. Vol. 108, No. 1, January 2000.

Wilhoit 1998. "A Computer Program to Identify Outliers in the Pesticide Use Report Database." April 1998. PM 98-01

Wilhoit 2002. "Pesticide Use Report Loading and Error-Handling Processes." January 2002. PM2-01

Wilhoit 2005. "PUR Update and Error Checking Proposal" April 2005. Slide presentation. David Supkoff is a co-author or co-presenter

APPENDIX A: DUPLICATE USAGE RECORDS ASSESSMENT

Corrections to Usage Data Measurement Units

For the exposure calculations the only variables reported in the PUR database that were required were: pounds of MeBr used, date, and location. However, in order to identify duplicate records for removal, as described below, additional variables were used. These included the areas treated and planted, which are recorded as the variables "ACRE_TREAT" and "ACRE_PLANT." These names are in some cases misnomers since the measurement units are not always acres. The measurement units are recorded in the variables "UNIT_TREAT" and "UNIT_PLANT." Possible values for MeBr are "A" = acres, "S" = square feet, and "U" = unspecified, typically numbers of plants or trees. Most of the MeBr data have Acres as the measurement units.

Our review of the CDPR usage data suggested that some of the values were erroneously reported as being measured in square feet when the actual measurement units were acres. In cases where UNIT_TREAT = "S" or "U" and UNIT_PLANT = "A," if ACRE_TREAT <= ACRE_PLANT, we corrected the treatment area units to UNIT_TREAT="A." In cases where UNIT_PLANT = "S" and UNIT_TREAT = "S," if ACRE_TREAT <= ACRE_PLANT and if the pounds of MeBr applied per acre >= 100,000 , we corrected the area units to UNIT_TREAT = "A" and UNIT_PLANT = "A."

Methods

Only those usage records with 400 or more pounds of MeBr were considered for the potential duplicate analysis. There were 26,185 such records between 1995 and 2001.

Two records were considered a match if all of the following held:

- Application dates are within 7 days
- Units treated match exactly
- "Acres" treated match within 5 % (i.e., the difference is less than 5 % of the average acres treated)
- Units planted match exactly
- "Acres" planted match exactly
- Grower ID numbers match exactly
- MTRS section numbers match exactly
- Pounds of product[28] used match within 1 % (i.e., the difference is less than 1 % of the average pounds of product used)
- Site Location ID values match exactly

[28] "Products" are different chemical mixture formulations containing various percentages of methyl bromide.

These pairs of matched records were joined into groups of 2 or more records by repeatedly using the rule that if records *a* and *b* form a matched pair, and if records *b* and *c* form a matched pair, then records *a*, *b*, and *c* are all in the same group.[29]

Generally, before 2001, growers were expected to make a single report of the total usage dated on the final application day. However, in some cases the growers did not adhere to these reporting procedures, and instead submitted a separate report for each application day. In order to avoid treating such multi-day application reports as duplicates two additional criteria were applied to the matched records.

- All records in the group are for the same application date, and the group total acres treated exceeds the acres planted (for each record) or
- Not all records in the group are for the same application date, and the group total acres treated exceeds 1.25 times the acres planted (for each record)

The 25 % adjustment factor accounts for the fact that the acres planted are often not reported as accurately as the acres treated.

We treated cases like the two Tulare County records mentioned above using an additional criterion: If the group has only two records and the average pounds of MeBr divided by the average acres treated is less than 200 and the units treated is acres, we assume that this unusually low application rate represents a field that was treated twice using half the normal application rate. Such groups are not treated as a group of duplicates.

From each remaining group of duplicates, all records except for the last one (i.e., the latest date) were deleted from the usage database.

Results

The attached file *duplicates.final.xls* lists all the duplicate records in each group including the record that was retained. The attached file *summary.of.duplicates.final.xls* summarizes the numbers of extra records that were removed, the pounds of MeBr used (group average), and the year. For example, if a group has three duplicate records, two records were removed. A total of 198 records were removed from 124 groups of duplicates.

[29] In some cases, because the matching criterion are not always exact, *a* and *c* may not form a match even if *a* matches *b* and *b* matches *c*.

APPENDIX B USAGE OUTLIER ASSESSMENT

We obtained a detailed list of outlier flags copied by Pesticide Action Network-North America (PANNA) from the PUR database (personal communications Jan 10, 2003, May 5, 2003, June 23, 2003, Oct 3 2003).. CDPR used three application rate outlier criteria and defined a record to be an outlier if any of these three criteria were met. There were 45,822 MeBr agricultural usage records in our database (from 1995 to 2001), of which 198 were removed as duplicate records as described above, leaving 45,624 unduplicated records. CDPR reported a total of 456 outliers (1.0 %), i.e., 456 records met at least one of their three outlier criteria. The CDPR analyses were based on each current year of data only.

For each criterion we present the CDPR approach and the alternative approach used for our analyses. Also discussed is the adjustment and correction of outlier records to the median application rate.

The criteria used or proposed by CDPR are described in Wilhoit (1998), Cal-EPA/CDPR (2000), Wilhoit (2002), and Wilhoit (2005). Wilhoit (1998) presents 5 initially proposed criteria. Cal-EPA/CDPR (2000) presents the three criteria that were used by CDPR to identify the outliers in the 1995-2001 usage data used for our analyses. Exactly three outlier criteria flags were included in the original database. Wilhoit (2005) describes the same three criteria and also refers to a fourth criteria (median + 50 times mean absolute deviation as being a current criterion. This is inconsistent with the PUR online documentation that describes the Outlier definition as being based on three criteria. The current system flags records as being an outlier based on all the outlier criteria being used, without indicating which criterion or criteria applied. Wilhoit (2005) also presents some proposed alternative approaches.

This document presents only the criteria for individual fumigants (such as methyl bromide) and for agricultural record types only.

Criteria

1. CDPR and ICF: More than 1000 pounds of MeBr used per acre.

This is Criterion 1a of Wilhoit (1998), Criterion 1 of Cal-EPA/CDPR (2000), and Criterion 1 of Wilhoit (2005). Wilhoit (1998) also evaluated an alternative criterion (1b) using a threshold of 2000 pounds per acre instead of 1000 pounds per acre.

CDPR only applied this test to records where the unit treated was acres.[30] The record meets this criteria if the pounds of MeBr used divided by the acres treated is greater than or equal to 1000.

We repeated the same procedure exactly. CDPR found 121 records meeting Criterion 1; we found 119 records meeting this criterion.[31]

[30] Excluding the cases where the unit of measurement is not acres (e.g., square feet) is not likely to change the overall usage volume, since these usages are very small.

[31] The two records flagged by CDPR but not ICF were:

2. CDPR: More than 50 times the median product used per unit treated for the same commodity, product, and measurement unit.

2. ICF: More than the mean plus three standard deviation of the logarithm[32] of MeBr used per unit treated for the same commodity, product, and measurement unit.

This is Criterion 2b of Wilhoit (1998), Criterion 2 of Cal-EPA/CDPR (2000), and Criterion 2 of Wilhoit (2005). Wilhoit (1998) also evaluated an alternative criterion (2a) using a threshold of 25 times the median instead of 50 times the median. In Wilhoit (1998) the matching was also by record-type, which is a more detailed classification than the Agricultural/Non-agricultural matching category that was used for the Cal-EPA/CDPR (2000) and Wilhoit (2005) versions of this criterion. CDPR used only data for a single year for these calculations. ICF used a broader approach, as follows.

In order to obtain sufficiently large sample sizes (i.e., at least 50 records) for making reasonably accurate approximations of the true mean and standard deviation, we adopted the following approach for grouping records.

- Case 1: Group the previous (if 1996 or later) and current year. No grouping of commodities[33] or products.
- Case 2: Group the previous (if 1996 or later) and current year. Group commodities where all but the last three digits of the site code[34] match. No grouping of products.
- Case 3: Group the previous (if 1996 or later) and current year. Group all products used. No grouping of commodities. Only applies if the area treated is reported as acres or square feet.
- Case 4: Group the previous (if 1996 or later) and current year. Group all products used. Group commodities where all but the last three digits of the site code match. Only applies if the area treated is reported as acres or square feet.
- Case 5: Group all years. No grouping of commodities or products.

- Yearuse_no 981862722, June 1998, 61.25 pounds MeBr used, 0.13 acres treated, 471 pounds per acre. CDPR records this as an outlier by this criterion, but we did not consider it as an outlier.
- Yearuse_no 982267262, October 1998, 396 pounds MeBr used, 1 acres treated, 396 pounds per acre. CDPR records this as an outlier by this criteria, but we did not consider it as an outlier based on the recorded pounds per acre..

[32] The logarithm of the application rate (pounds of product used per unit treated) was used for these calculations, since CDPR noted that the distribution of the application rate is skewed. This suggests that a log transformation will make the data more symmetric and closer to a normal distribution.

[33] "Commodities" are the agricultural products such as fruits in general or specific fruits such as strawberries

[34] The site code indicates the "commodity". The site (commodity) codes used in the PUR database are listed in the attached Excel file *Site.xls*.

- Case 6: Group all years. Group commodities where all but the last three digits of the site code match. No grouping of products.

For each usage record, we applied each case in turn until we obtained a sample of at least 50 from which to calculate a mean and standard deviation. For example, if cases 1 and 2 have less than 50 records but case 3 has 50 records or more, then case 3 is applied.

If the units treated is measured in square feet or acres, and if the logarithm of the application rate in pounds MeBr used per acre treated exceeds the mean plus three standard deviations of the logarithm of the application rate in pounds MeBr per acre treated for the appropriate case, then the given record is considered a usage outlier.

If the units treated is not measured in square feet or acres, and if the logarithm of the application rate in pounds product used per unit treated exceeds the mean plus three standard deviations of the logarithm of the application rate in pounds product used per unit treated for the appropriate case (1, 2, 5 or 6), then the given record is considered a usage outlier.

If all cases have less than 50 records, but CDPR found this record to be an outlier by their Criterion 2, and if the pounds MeBr used per acre exceeds 480 (i.e., more 20% above the legal limit of 400 lbs/acre), then we also considered the record an outlier.[35]

If none of these cases applied, then the corresponding records were not considered as outliers.

CDPR found 329 records meeting their version of Criterion 2, but we found 212 records meeting this criterion (44 records were usage outliers by both methods).

3. CDPR and ICF: Neural network.

This is Criterion 4 of Wilhoit (1998), Criterion 3 of Cal-EPA/CDPR (2000), and Criterion 4 of Wilhoit (2005). In Wilhoit (1998) the matching was also by record-type, which is a more detailed classification than the Agricultural/Non-agricultural matching category that was also included in the CDPR (2000) and Wilhoit (2005) versions of this criterion. CDPR used only data for a single year for these calculations.

A neural network procedure maps the summary statistics (e.g. mean, median) of the application rates for the current year, commodity, product, and measurement unit into an upper threshold value. The mapping was based on a review by 12 CDPR scientists of "typical" distributions of application rates (normalized to a median of 1) and their determinations of the upper thresholds for each of those distributions. These reviews were used as a "training" data set for the neural network procedure. After reviewing the outlier flags provided by CDPR we found some inconsistencies. We discussed these inconsistencies with Larry Wilhoit of CDPR who discovered a software problem that had caused many of the Criterion 3 outlier flags to have been

[35] These extra cases found by CDPR are suspected of not following the log-normality assumed in our general approach to Criterion 2

wrongly set. Mr. Wilhoit then provided us the attached file *corrupted _nn.xls* that lists all the corrupted records (for all chemicals, not just MeBr).

In the absence of detailed documentation, it was not feasible for us to revise or update this procedure. A total of 113 agricultural usage records (0.25 %) were flagged by this criterion. CDPR had developed the thresholds based only on the data in the current year. But for this analysis we applied this criterion only if there was evidence of some stability in the year to year distributions. Criterion 3 outliers were retained as outliers if the median application rates for the commodity, product, and measurement unit in the current and previous years were within a factor of 10, i.e. the ratio was between 0.1 and 10, inclusive. Records with higher or lower ratios were not retained as outliers for this analysis. CDPR Criterion 3 outlier records with no values for the previous year were also retained as outliers.

CDPR found 113 records meeting their version of Criterion 3, but we found 78 records meeting the revised criterion, i.e., 78 of those 113 records had current and previous year median application rates that differed by an order of magnitude (factor of 10).

4. CDPR: More than the sum of the median and 50 times the median absolute deviation (MAD) of the product used per acre treated for the same commodity, product, and measurement unit.

This is Criterion 3b of Wilhoit (1998) and Criterion 3 of Wilhoit (2005). The median absolute deviation is defined as the median of the absolute differences between an observed value and the median observed value; this is a non-parametric statistic that measures variability analogously to the standard deviation measure of variability for the normal and other parametric distributions.

This criterion is not listed in Cal-EPA/CDPR (2000) and was not used by ICF. In Wilhoit (1998) the matching was also by record-type, which is a more detailed classification than the Agricultural/Non-agricultural matching category that was also included in the Wilhoit 2005 versions of this criterion. CDPR used only data for a single year for these calculations. Wilhoit 1998 also proposed a version 3a of this criterion using a threshold of median + 10× MAD instead of median + 50 × MAD. One slide in Wilhoit 2005 presented this option as median + 20 × MAD.

5. CDPR: Acres treated greater than 700.

This criterion is Criterion 5 of Wilhoit (1998). Although not included in Cal-EPA/CDPR (2000) and Wilhoit (2005), the report by Wilhoit (2002) implies that the PUR data loading program automatically checks for this condition and flags this as an error. Thus this criterion is automatically applied when the data are entered.

Proposed Criteria

Wilhoit (2005) also made some proposals for alternative criteria. He proposes using a 5-year period to identify outlires and also proposed using the logarithms of the rates instead of the rates (similarly to ICF's Criterion 2). Additional proposed criteria are as follows:

Product used per acre treated for the same commodity, product, and measurement unit is at least:

- 50 × Median by AI
- Median + 5 × STD MAD
- Trimmed mean + 5 × Trimmed MAD
- 75^{th} percentile + 3 × IQR
- 75^{th} percentile + 3 × STD IQR

In the absence of more detailed documentation it is not at all clear how these statistics are defined and calculated and how well these alternative criteria would perform compared to the other criteria.

Final Results

Out of the 45,624 non-duplicate agricultural records for 1995-2001, CDPR reported 456 (1.0 %) outlier records, i.e., records that met at least one of the three criteria. Our alternative approach produced a total of 329 (0.72 %) outlier records that met at least one of our three criteria.[36] Of these 329 records, 207 were also outliers according to CDPR, but 122 of these 329 records were outliers according to our criteria but not according to CDPR's criteria, i.e., they met our Criterion 2 and not CDPR's.

These 329 outlier records are listed in the attached Excel spreadsheet *outliers.final.xls,* in the worksheet "ICF". The worksheet "DPR only" lists the 249 records that were outliers based on CDPR's methodology but not outliers based on our methodology. The columns in the spreadsheet display the detailed results of the outlier analyses.

Adjustment of Outliers

As suggested by CDPR's initial treatment of suspected outliers, any records found to be outliers were adjusted by reducing the application rate to the median level. Assume first that the measurement unit for the area treated was acres or square feet. The data were converted into acres. If the pounds MeBr used per acre were less than 480 (20 % above the legal limit, to allow for uncertainties in the reported acres treated), then no adjustment was made. Otherwise, for consistency with our Criterion 2, the median application rate (pounds MeBr applied per unit treated) was computed for cases 1, 2, 3, 4, 5, and 6 defined above. Case 1 is for the given product, site code and measurement using the median over the current and previous years. Cases

[36] Of interest is the consistency between the treatments of outliers for the monitoring study regressions (i.e., regression, model calibration and selection) and for this statewide application. One record for the monitoring period met our outlier Criterion 2, and was adjusted for use in this statewide application as explained below, but was used without adjustment in the regression modeling. This is the record with yearuse_no 01603683:

- 01603686: 27 September 2001, MTRS M13S02E24, Monterey, reported 2456.13 pounds of MeBr, Adjusted value (see below) 579.94 pounds of MeBr.

2, 3, 4, 5, and 6 combine records by commodity group (all but the last three digits of the site code) and/or product and/or years. The median used for adjustment was from the first of these cases with 50 or more records. If there were no cases with 50 or more records, then the case with the most records was used. The adjusted amount of MeBr used equals the acres treated multiplied by the application rate (pounds MeBr applied per acre treated).

If the measurement unit for the area treated was not "A" or "S," then the median application rate in pounds of product per unit treated was calculated for cases 1, 2, 5, and 6 only. The median used for adjustment was from the first of these cases with 50 or more records. If there were no cases with 50 or more records, then the case with the most records was used. The adjusted amount of MeBr used equals the area treated ("ACRE_TREAT") multiplied by the application rate in pounds product applied per unit treated and multiplied by the MeBr fraction (pounds MeBr per pounds of product).

The results of these usage adjustments are given in the worksheet "ICF" in the *outliers.final.xls* file. The original and adjusted amounts of MeBr used are shown for each outlier record along with the difference and percentage difference.

References

Cal-EPA/CDPR 2000. *Pesticide Use Reporting; an Overview of California's Unique Full Reporting System*. May 2000. http://www.cdpr.ca.gov/docs/pur/purovrvw/ovr52000.pdf

Cal-EPA/CDPR 2002. *Pesticide Use Report Data: User Guide & Documentation* Department of Pesticide Regulation, Information Systems Division. http://www.krisweb.com/biblio/cal_dpr_dpr_2002_pesticidecdguide.pdf

Personal Communication, January 10, 2003, Susan Kegley, Pesticide Action Network, Provided: California 2001 Methyl Bromide Usage Data by Meridian Township Section Range.

Personal Communication, May 5, 2003, Susan Kegley, Pesticide Action Network, Provided: California 1995 Methyl Bromide Usage Data by Meridian Township Section Range.

Personal Communication, June 23, 2003, Stephan Orme, Pesticide Action Network, Provided: California 1996-2000 Methyl Bromide Usage Data by Meridian Township Section Range.

Personal Communication, October 3, 2005, Susan Kegley, Pesticide Action Network, Provided: California 2002 and 2003 Methyl Bromide Usage Data by Meridian Township Section Range,

Wilhoit 1998. "A Computer Program to Identify Outliers in the Pesticide Use Report Database." April 1998. PM 98-01

Wilhoit 2002. "Pesticide Use Report Loading and Error-Handling Processes." January 2002. PM2-01

Wilhoit 2005. “PUR Update and Error Checking Proposal” April 2005. Slide presentation. David Supkoff is a co-author or co-presenter

APPENDIX C. MULTI-DAY APPLICATIONS

Generally, up to the year 2001, growers were expected to make a single report of the total usage dated at the final application day rather than reporting each daily application.. This Appendix describes the approach used for determining which records are likely to refer to a multi-day application and for allocating the reported MeBr usage for such records across multiple days over the time period of study. This approach was based on conversations with Paul Niday of Tri-Cal applicators, one of the largest MeBr applicators in the State, and three county agricultural commissioner officers in three counties, Ventura, Tulare and Riverside, where large MeBr applications have historically occurred.

The approach is divided into two time periods: those applications which occurred prior to the revised State MeBr application regulations which became effective on 14 January 2001, and the applications following these new procedures. The new regulations decreased the maximum allowable application rate and increased the minimum waiting time between applications at a single location. These regulations are fully described in *Title 3 of the California Code of Regulation, Division 6. Pesticides and Pest Control Operations, Chapter 2. Pesticides, Subchapter 4. Restricted Materials, Article 4. Use Requirements, Section 6450. Chloropicrin and Methyl Bromide Field Fumigation.*

Exclusions

As described below, the multi-day algorithm temporally apportions usages above 40 acres prior to 14 January, 2001 and usages above 30 acres on or after 14 January, 2001 to multiple days. In applying this algorithm we first excluded all usages with less than 400 pounds of MeBr usage from consideration (i.e., we assumed that the usage occurred on a single day). The county agricultural commissioner officers and Tri-Cal staff indicated that after the main MeBr application, areas are sometimes spot treated so that the MeBr is applied to a few locations inside a much larger area. These spot treatments can in most cases be identified in the database as having large treatment areas and small application rates. The choice of 400 pounds of MeBr as an upper bound for exclusion means that the application rate for excluded records would be at most 400/40 = 10 pounds per acre prior to 14 January, 2001 and at most 400/30 = 13.3 pounds per acre on or after 14 January, 2001. MeBr application rates are typically in the 100 to 400 pounds per acre range, depending upon the commodity.

Usage records that were included in the regression analysis were also excluded from consideration for the multi-day algorithm[37]. This is because we previously examined all of these records with more than 100 pounds per acre usage, found which ones were multi-day applications, and apportioned the usage to the exact usage days. These same temporal allocations were used for this statewide application instead of applying the general rules described below[38].

[37] The regression analysis could potentially have used any records in Kern County from 5 July, 2000 to 1 September, 2000 and from 14 June, 2001 to 30 August, 2001 and records in Monterey, Santa Cruz, San Benito, and Santa Clara counties from 28 August, 2000 to 2 November, 2001 and from 27 August, 2001 to 7 November, 2001.

[38] For the counties and time periods listed in the footnote above, there were also two usage records that had more than 400 pounds of methyl bromide but less than 100 pounds per acre of usage. Those two records were assumed not

Usage Adjustments

Case 1: Prior to 14 January 2001

Standard practice prior to the new regulations was to operate 2 to 4 MeBr application tractor teams on the same field. Typically up to 10 acres per day could be completed by each tractor team.

If we assume that in most of these applications that 4 tractor teams are used, the number of application days can be expressed as:

$$\text{Number of application days} = (\text{acres treated})/(4 \times 10 \text{ acres/day}) \qquad \text{(C-1)}$$

So, for example, a 130-acre treatment would take: $130/(4 \times 10) = 3.25$ application days

We therefore chose to temporally apportion any usage for each reported fumigation event covering more than 40 acres by:

1. Calculating the number of application days according to equation C-1;
2. Rounding the number of days up to the next integer, *N*; and
3. Allocating the usage equally across days *X*-(*N*-1), *X*-(*N*-2), … *X*, where *X* is the recorded date of application.

The usage for each day would be the total MeBr usage reported divided by *N*.

In the example the 130-acre treatment would be apportioned evenly across 4 days, including the reported application date and the three days immediately preceding the reported application date. The amount of MeBr usage for each day would be:

$$\text{MeBr per day} = (\text{total MeBr reported})/4$$

Case 2: Post 14 January 2001

Because of additional requirements in application procedures under the new regulations typically only 7.5 acres can be completed each day by a tractor team. Again, for larger acreage 2 to 4 tractor teams may operate on a field. Further, the regulations only permit a maximum of 40 acres to be treated in a single day, and require a 36-hour waiting period between applications.

If we assume that for most of these applications 4 tractor teams are used, the number of application days can be expressed as:

$$\text{Number of application days} = NAD = (\text{acres treated})/(4 \times 7.5 \text{ acres/day}) \qquad \text{(C-2)}$$

---

to be multi-day applications but to be spot treatments, since there were recent previous records for the same grower and field identifier (SITE_LOC_ID).

The total number of application days would then be increased for the 36-hour waiting period using the following expression:

$$\text{Duration of application} = D = (\text{roundup}(NAD) \times 2) - 1, \qquad \text{(C-3)}$$

where roundup(*NAD*) = *NAD* rounded up to the nearest integer.

So for example, a 130-acre treatment would take:

$130/(4 \times 7.5)$ = 4.3 application days

$(\text{roundup}(4.3) \times 2) - 1$ = 9 day duration period

We temporally apportioned any usage for each application to more than 30 acres by:

1. Calculating the number of application days according to equation C-2;
2. Rounding the number of days up to the next integer, *N*;
3. Calculating the duration of application, *D*, according to equation C-3;
4. Allocating the usage equally across days *X*-(*D*-1), *X*-(*D*-3), … *X*, where *X* is the recorded date of application.

The usage for each day would be the total MeBr usage reported divided by *N*.

Thus for Case 2, the usage would be temporally apportioned using the same approach as for Case 1, except that the treatment would be assumed to occur on every second day ending on day *X*.

APPENDIX D. FMMP METADATA

fmmp_meta.txt 14.mar.01

This file includes metadata information pertinent to both the FMMP-native Microstation design file (dgn) format and to the ArcShape export format.

State of California
Department of Conservation
Division of Land Resource Protection
FARMLAND MAPPING AND MONITORING PROGRAM

Coverage: Important Farmland (agricultural land use)
Location: ftp.consrv.ca.gov/pub/dlrp/fmmp/
Contact:
Division of Land Resource Protection
Farmland Mapping and Monitoring Program
801 K Street, MS 13-71
Sacramento, CA 95814
(916) 324-0859
http://www.consrv.ca.gov/dlrp/fmmp
fmmp@consrv.ca.gov

General Description
The FMMP compiles two kinds of land use maps:
Important Farmland Maps for those areas that have modern soil surveys, and Interim Farmland Maps for those areas lacking modern soil survey information and for which there is expressed local concern on the status of farmland. (Technical definitions for these map categories can be found in "A Guide for the Farmland Mapping and Monitoring Program".)

Important Farmland Map Categories
Important Farmland maps for California are compiled using USDA-NRCS Soil Surveys and current land use information using eight mapping categories generally explained below. The minimum mapping unit is 10 acres. Units of land smaller than 10 acres are incorporated into the surrounding map categories.

Prime Farmland (P)
Irrigated land with the best combination of physical and chemical features able to sustain long term production of agricultural crops. This land has the soil quality, growing season, and moisture supply needed to produce sustained high yields. Land must have been used for production of irrigated crops at some time during the four years prior to the mapping date.

Farmland of Statewide Importance (S)
Irrigated land similar to Prime Farmland that has a good combination of physical and chemical characteristics for the production of agricultural crops. This land has minor shortcomings, such as greater slopes or less ability to store soil moisture than Prime Farmland. Land must have been

used for production of irrigated crops at some time during the four years prior to the mapping date.

Unique Farmland (U)
Lesser quality soils used for the production of the state's leading agricultural crops. This land is usually irrigated, but may include non-irrigated orchards or vineyards as found in some climatic zones in California. Land must have been cropped at some time during the four years prior to the mapping date.

Farmland of Local Importance (L and LP)
Land of importance to the local agricultural economy as determined by each county's board of supervisors and a local advisory committee. See either "A Guide to the Farmland Mapping and Monitoring Program", or the latest copy of the "Farmland Conversion Report" for each county's definition of Farmland of Local Importance.

Grazing Land (G)
Land on which the existing vegetation is suited to the grazing of livestock. This category is used only in California and was developed in cooperation with the California Cattlemen's Association, University of California Cooperative Extension, and other groups interested in the extent of grazing activities. The minimum mapping unit for Grazing Land is 40 acres.

Urban and Built Up Land (D)
Land occupied by structures with a building density of at least 1 unit to 1.5 acres, or approximately 6 structures to a 10-acre parcel.

Other Land (X)
Land which does not meet the criteria of any other category. Typical uses include low density rural development, heavily forested land, mined land, or government land with restrictions on use.

Water (W)
Water areas with an extent of at least 40 acres.

Area Not Mapped (Z)
Area which falls outside of the NRCS soil survey. Not mapped by the FMMP.

Interim Farmland Map Categories
Interim Farmland maps are prepared for specific agricultural counties lacking modern soil surveys. Two categories of Interim Farmland are mapped in lieu of Prime Farmland, Farmland of Statewide Importance, Unique Farmland, and Farmland of Local Importance. The categories of Irrigated Farmland and Non-Irrigated Farmland are designed to be easily upgraded as the USDA-NRCS completes modern soil survey mapping as the technical soil information becomes available. The Interim Farmland maps also utilize the categories of Grazing Land, Urban and Built-up Land, Other Land, and Water as defined under the section 'Important Farmland Map Categories'.

Irrigated Farmland (I)
Cropped land with a developed irrigation water supply that is dependable and of adequate quality. Land must have been used for production of irrigated crops at some time during the four years prior to the mapping date.

Non-irrigated Farmland (N)
Land on which agricultural commodities are produced on a continuing or cyclic basis utilizing stored soil moisture.

Draft Map Categories
Early versions of maps may have the categories of iP and nG, which are mixed irrigated pasture and dry grain, respectively. These land uses are often incorporated into the Farmland of Local Importance definitions for counties once the initial map has undergone review by a Local Advisory Committee and the County Board of Supervisors.

Vital Statistics
Projection: Albers Equal Area
Datum: NAD 27
Ellipsoid: Clarke 1866
Longitude of Origin: -120:00:00
Latitude of Origin: 0:00:00
Standard Parallel 1: 34:00:00
Standard Parallel 2: 40:30:00
False Easting: 0.00 meters
False Northing: -4000000.000 meters
Working Units (DGN format): 1,000 UORs per meter
Storage Center Point (DGN format): X: 0.00 meters, Y: 0.00 meters

Source: California Department of Conservation
Farmland Mapping and Monitoring Program
Source Media: Aerial photographs of various scale (Mostly 1:130,000 scale CIR) and field reconnaissance.
Also, 1:24,000 scale USGS topographic quadrangles
Source Projection: Mostly conformal conic
Source Units: N/A
Source Scale: 1:24,000 AND 1:100,000
Capture Method: Digitized
Data Structure: Vector
Data Coverage: County
Precision Floating point
Average File Size: 1mb
Updated: Biennially

Shape File Format:
Relevant fields included in Farmland Mapping and Monitoring Program data:

polygon_ty - abbreviation of Important Farmland categories. For translation, see key below.
polygon_ac - acreage of polygons

polygon_ty	longname
P	Prime Farmland
S	Farmland of Statewide Importance
U	Unique Farmland
L	Farmland of Local Importance
LP	Farmland of Local Potential
I	Irrigated Farmland
N	Non-Irrigated Farmland
iP	Irrigated Pasture
nG	Non-Irrigated Grain
G	Grazing Land
D	Urban and Built-Up Land
X	Other Land
W	Water
Z	Not Inventoried

Design (Dgn) File Format:
Level 1 Important Farmland linework (color=2 or 3)
Level 2 centroids (color=4)
Level 5 mapping category text (see map category reference below)
Level 58 Caveats and restrictions
Level 61 FMMP ID cells
Level 62 USGS 7.5 minute quadrangle names
Level 63 7.5 minute tics

Level 5 category text can be cross referenced using the key below:

P	Prime Farmland
S	Farmland of Statewide Importance
U	Unique Farmland
L	Farmland of Local Importance
LP	Farmland of Local Potential
I	Irrigated Farmland
N	Non-Irrigated Farmland
iP	Irrigated Pasture
nG	Non-Irrigated Grain
G	Grazing Land

D	Urban and Built-Up Land
X	Other Land
W	Water
Z	Not Inventoried

Copyright And Use Restrictions

The following information applies to all of the maps produced by the Farmland Mapping and Monitoring Program: © Department of Conservation, Farmland Mapping and Monitoring Program, 2001. The Department of Conservation makes no warranties as to the suitability of this product for any particular purpose. Use of these data at any scale larger than 1:100,000 is considered an enlargement.

Data Quality Assessment

The following are subjective comments regarding this data.

The extent of the farmland coverage corresponds to the availability of what the National Resources Conservation Service (NRCS) considers "modern soil surveys". Thus, coverage is so far limited to 48 of California's 58 counties. Some of these 48 counties have only been partially mapped as the NRCS has limited its efforts to areas that are agriculturally important. The soil surveys rarely extend into National Forests.

Statistical note:

Text files associated with this site that provide acreage figures for quality assurance purposes more accurately affect the files on the ftp site than the figures in older Farmland Conversion Reports. Changes made to files (projection change, standardized county line file, ellipsoidal calculations) since the publication of the Conversion Reports are not reflected in the published documents. Use the Text files and/or Excel files on the FMMP web site for comparisons.

APPENDIX E. LUSD MATADATA

Below is sample metadata for the LUSD dataset for Colusa County. A separate metadata file is available at the provided website for each county.

Metadata for the 1998 Colusa County Land Use Survey Data

Originator: California Department of Water Resources

Date of Metadata: May 14, 2001

Abstract:
The 1998 Colusa County land use survey data set was developed by DWR through its Division of Planning and Local Assistance. The data was gathered using aerial photography and extensive field visits, the land use boundaries and attributes were digitized, and the resultant data went through standard quality control procedures before finalizing. The land uses that were gathered were detailed agricultural land uses, and less detailed urban and native vegetation land uses. The data was gathered and digitized by staff of DWR's Northern District and the quality control procedures were performed jointly by staff at DWR's DPLA headquarters from San Joaquin District. The finalized data include DWG files (land use vector data), shapefiles (land use vector data), and JPEG files (raster data from aerial imagery).

Purpose:
This data was developed to aid in DWR's efforts to continually monitor land use for the main purpose of determining the amount of and changes in the use of water.

DWR Contacts:

Tito Cervantes
Northern District
2440 Main Street
Red Bluff, CA 96080-2356
530-529-7389
cervante@water.ca.gov

Tom Hawkins
DPLA Headquarters
1416 9th Street
Sacramento, CA 95814
916-653-5573
hawkins@water.ca.gov

Data Development:

1. The aerial photography used for this survey was taken in late June of 1998. The photos (natural color, 9" by 9", flown at 18,000' above ground with a 6" lens) were scanned at 300 DPI and plotted to a size of about 20" x 20".

2. The plotted images were taken to the field as field sheets, and virtually all the areas were visited to positively identify the land use. The site visits occurred in July through September 1998. Land use codes were printed within each area on the field sheets.

3. For those areas where the elevation changes were minimal, the scanned images were brought into an image processing system, the images were ratio-rectified (rubbersheeted) into a projection and mosaiced into USGS 1:24,000 quad sized files (photoquads). The files have a pixel size of 4 meters.

4. Using AUTOCAD (using a standardized digitizing process), the photoquads were used as a backdrop to delineate land use boundaries on-screen. For those areas where corrected imagery was not produced (because of excess elevation changes), land use boundaries were drawn onto USGS 1:24,000 quads, and those quad maps digitized on a digitizing tablet. The land use attributes were entered from the field sheets.

5. After quality control/assurance procedures were completed on each file (DWG), the data was finalized.

6. The linework and attributes from each DWG quad file were brought into ARCINFO and both quad and surveywide coverages were created, and underwent quality checks. These coverages were converted to shape files using ARCVIEW.

Data Accuracy:
Linework for those areas where photoquads were developed: The land use boundaries were drawn on-screen in AUTOCAD using the photoquads as a backdrop. The resultant digital linework for those areas is at best 100 foot accuracy. Linework for those areas where photoquads were not developed: The land use boundaries were hand drawn onto USGS 1:24,000 quads, and digitized on a digitizing tablet using AUTOCAD. For those areas where the lines were drawn onto USGS quads and digitized, the accuracy is less than that of the quads (about 50 foot accuracy). The land use attribute accuracy is very high, because almost every delineated field was visited in the field. The accuracy is less than 100 percent because some errors must have occurred. There are three possible sources of attribute errors that are:

(1) Misidentification of land use in the field (and entering that incorrect attribute on the field sheet);

(2) Correct identification of land use, but entering an incorrect attribute on the field sheet, or;

(3) Accidentally affixing an incorrect attribute during the digitizing process.

The corrected imagery (photoquads) was developed using between 12 and 15 ground control points established from terrain corrected satellite imagery with a stated accuracy of about 30 feet. The imagery has never been fully evaluated for positional accuracy, however we believe that the images have about 100 foot accuracy (90 percent of the time, the data is within 100 feet of its true position).

Projection Information:
The data (DWG, shape files, and corrected imagery) is in a transverse mercator projection, with identical parameters to UTM projections, except the central meridian is -120 degrees (120 degrees west). For comparison, UTM 10 has a central meridian of 123 degrees west, and UTM 11 has a central meridian of 117 degrees west. This projection allows virtually all of the geographic area of California to be in one 6 degree zone (as opposed to two zones, UTM 10 and 11).

Projection:	Transverse Mercator
Datum:	NAD27
Units:	Meter
Scale Reduction:	0.9996
Central Meridian:	120 degrees west
Origin Latitude:	0.00 N
False Easting:	500,000
False Northing:	0.00

Land Use Attributes:
All land use attributes were coded using the Department's Standard Land Use Legend dated March 1999 (98legend.pdf). The legend explains in detail how each delineated area is attributed in the field, and what the coding system is. The actual land use code that is printed onto the field maps is different in arrangement than the codes that result from the digitizing process. The file attributes.pdf is a detailed explanation of the coding system used for both coding the field sheets, and the codes that end up in digitized form in the database files associated with the shape files.

Information on the AUTOCAD (DWG) Files:
The land use data is available in AUTOCAD 12 format by quad, with one file per quad. The file naming convention is 98COXXXX.DWG, where XXXX is the DWR quadrangle number. For example, file 98CO2320.DWG is the AUTOCAD drawing file for the 1998 Colusa County land use survey for quadrangle 2320 (the Colusa quad). Every quadrangle file has identical layers, nomenclature, and line colors. They are as follows:

Layer	Description	Color
0	AutoCAD's default layer	White
CQN	California DWR quad number	Cyan
GSN	USGS quad number	Cyan
LUB	Land use boundary lines	Yellow
LUC	Land use codes for GRASS	White
LUT	Visible land use text	Green
QB	The quad's boundary	White
QN	Quad name	Cyan

Following is an explanation of the attributes (for each delineated area) in the LUC layer of each quad file:

ACRES:	Number of acres in the delineated area (may or may not be present)

WATERSOURC:	The type of water source used for the delineated area
MULTIUSE:	Type of land uses within the delineated area
CLASS1:	The class for the first land use
SUBCLASS1:	The subclass for the first land use
SPECOND1:	The special condition for the first land use
IRR_TYP1:	Irrigated or non-irrigated, and irrigation system type for the first land use
PCNT1:	The percentage of land associated with the first land use
CLASS2:	The class for the second land use
SUBCLASS2:	The subclass for the second land use
SPECOND2:	The special condition for the second land use
IRR_TYP2:	Irrigated or non-irrigated, and irrigation system type for the second land use
PCNT2:	The percentage of land associated with the second land use
CLASS3:	The class for the third land use
SUBCLASS3:	The subclass for the third land use
SPECOND3:	The special condition for the third land use
IRR_TYP3:	Irrigated or non-irrigated, and irrigation system type for the third land use
PCNT3:	The percentage of land associated with the third land use

Information on the Shape Files:
Shape files were created for each quad, and one for the whole survey area. The naming conventions used for the quad DWG files is used for the quad shape files (for example, 98CO2320.shp, 98CO2320.shx, and 98CO2320.dbf for quad number 2320, the Colusa quad). The name of the shape file for the whole survey area is 98CO.shp (and .dbf and .shx).

Following is an explanation of the land use attributes in the DBF files:

BL_X:	This is the X coordinate of the interior point in the delineated area
BL_Y:	This is the Y coordinate of the interior point in the delineated area
ACRES:	Number of acres in the delineated area (may or may not be present)
WATERSOURC:	The type of water source used for the delineated area
MULTIUSE:	Type of land uses within the delineated area
CLASS1:	The class for the first land use
SUBCLASS1:	The subclass for the first land use
SPECOND1:	The special condition for the first land use
IRR_TYP1A:	Irrigated or non-irrigated for the first land use
IRR_TYP1B:	Irrigation system type for the first land use
PCNT1:	The percentage of land associated with the first land use
CLASS2:	The class for the second land use
SUBCLASS2:	The subclass for the second land use
SPECOND2:	The special condition for the second land use
IRR_TYP2A:	Irrigated or non-irrigated for the second land use
IRR_TYP2B:	Irrigation system type for the second land use
PCNT2:	The percentage of land associated with the second land use
CLASS3:	The class for the third land use
SUBCLASS3:	The subclass for the third land use

SPECOND3:	The special condition for the third land use
IRR_TYP3A:	Irrigated or non-irrigated for the third land use
IRR_TYP3B:	Irrigation system type for the third land use
PCNT3:	The percentage of land associated with the third land use
UCF_ATT:	Concatenated attributes from MULTIUSE to PCNT3

Information on the JPEG Files:
JPEG files were created for each quad where there was a minimum of elevation changes. The file naming convention is 984CXXXX.JPG, where 98 is the year, 4C is for "Four Counties" (Glenn, Colusa, Sutter, and Yolo), and XXXX is the DWR quadrangle number. For example, files 984C2320.JPG and 984C2320.JGW are the quad files for the 1998 Four County land use survey for quadrangle 2320 (the Colusa quad). The .JGW file is the JPEG world file.

Important Points about Using this Data Set:

1. The land use boundaries were either drawn on-screen using developed photoquads, or hand drawn directly on USGS quad maps and then digitized. They were drawn to depict observable areas of the same land use. They were not drawn to represent legal parcel (ownership) boundaries, or meant to be used as parcel boundaries.

2. This survey was a "snapshot" in time. The indicated land use attributes of each delineated area (polygon) were based upon what the surveyor saw in the field at that time, and, to an extent possible, whatever additional information the aerial photography might provide. For example, the surveyor might have seen a cropped field in the photograph, and the field visit showed a field of corn, so the field was given a corn attribute. In another field, the photograph might have shown a crop that was golden in color (indicating grain prior to harvest), and the field visit showed newly planted corn. This field would be given an attribute showing a double crop, grain followed by corn. The DWR land use attribute structure allows for up to three attributes per delineated area (polygon). In the cases where there were crops grown before the survey took place, the surveyor may or may not have been able to detect them from the field or the photographs. For crops planted after the survey date, the surveyor could not account for these crops. Thus, although the data is very accurate for that point in time, it may not be an accurate determination of what was grown in the fields for the whole year. If the area being surveyed does have double or multicropping systems, it is likely that there are more crops grown than could be surveyed with a "snapshot".

3. If the data is to be brought into a GIS for analysis of cropped (or planted) acreage, two things must be understood:

    a. The acreage of each field delineated is the gross area of the field. The amount of actual planted and irrigated acreage will always be less than the gross acreage, because of ditches, farm roads, other roads, farmsteads, etc. Thus, a delineated corn field may have a GIS calculated acreage of 40 acres but will have a smaller cropped (or net) acreage, maybe 38 acres.

    b. Double and multicropping must be taken into account. A delineated field of 40 acres might have been cropped first with grain, then with corn, and coded as such. To

estimate actual cropped acres, the two crops are added together (38 acres of grain and 38 acres of corn) which results in a total of 76 acres of net crop (or planted) acres.

4. Water source and irrigation type information were collected for this survey, but are not present in this dataset. Contact Tito Cervantes of Northern District for more information about this data.

APPENDIX F. DETAILED REVIEW OF NLCD AGRICULTURAL DATA

In an effort to compile a single dataset of California farmlands, the NLCD satellite classification data was considered as a supplement to the California Department of Conservation and Department of Water Resources farmland data. NLCD was available for the entire country, had a methodology that was consistently applied and was thought to be accurate enough to supplement the other two data sources.

Background

A consortium consisting of several federal agencies was formalized to produce a consistent and seamless National Land Cover Data set (NLCD) for the conterminous United States. Land cover mapping has been conducted for each of ten geographic regions using early 1990s Landsat Thematic Mapper (TM) imagery augmented by a suite of other geospatial ancillary data sets. Briefly, the NLCD was compiled through unsupervised clustering of Landsat TM data. The resulting spectral clusters were resolved into one of 21 thematic classes using logical modeling and ancillary data sources (e.g., census, slope/aspect/elevation, etc.) as required. The twenty-one thematic classes resemble the well-established Anderson land use/cover classification system. Both leaves-off and leaves-on data sets were analyzed. The resulting clusters were then labeled using aerial photography and ground observations. Clusters that represented more than one land cover category were also identified and, using various ancillary data sets, models were developed to split the confused clusters into the correct land cover categories.

Categories of NLCD

Water - All areas of open water or permanent ice/snow cover.

11 - Open Water
All areas of open water, generally with less than 25% cover of vegetation/land cover.

12 - Perennial Ice/Snow
All areas characterized by year-long surface cover of ice and/or snow.

Developed Areas characterized by a high percentage (30 percent or greater) of constructed materials (e.g. asphalt, concrete, buildings, etc).

21 - Low Intensity Residential
Includes areas with a mixture of constructed materials and vegetation. Constructed materials account for 30-80 percent of the cover. Vegetation may account for 20 to 70 percent of the cover. These areas most commonly include single-family housing units. Population densities will be lower than in high intensity residential areas.

22 - High Intensity Residential

Includes highly developed areas where people reside in high numbers. Examples include apartment complexes and row houses. Vegetation accounts for less than 20 percent of the cover. Constructed materials account for 80 to100 percent of the cover.

23 - Commercial/Industrial/Transportation

Includes infrastructure (e.g. roads, railroads, etc.) and all highly developed areas not classified as High Intensity Residential.

Barren - Areas characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the "green" vegetated categories; lichen cover may be extensive.

31 - Bare Rock/Sand/Clay

Perennially barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, beaches, and other accumulations of earthen material.

32 - Quarries/Strip Mines/Gravel Pits

Areas of extractive mining activities with significant surface expression.

33 - Transitional

Areas of sparse vegetative cover (less than 25 percent of cover) that are dynamically changing from one land cover to another, often because of land use activities. Examples include forest clearcuts, a transition phase between forest and agricultural land, the temporary clearing of vegetation, and changes due to natural causes (e.g. fire, flood, etc.).

Forested Upland - Areas characterized by tree cover (natural or semi-natural woody vegetation, generally greater than 6 meters tall); tree canopy accounts for 25-100 percent of the cover.

41 - Deciduous Forest

Areas dominated by trees where 75 percent or more of the tree species shed foliage simultaneously in response to seasonal change.

42 - Evergreen Forest

Areas dominated by trees where 75 percent or more of the tree species `maintain their leaves all year. Canopy is never without green foliage.

43 - Mixed Forest

Areas dominated by trees where neither deciduous nor evergreen species represent more than 75 percent of the cover present.

Shrubland - Areas characterized by natural or semi-natural woody vegetation with aerial stems, generally less than 6 meters tall, with individuals or clumps not touching to interlocking. Both evergreen and deciduous species of true shrubs, young trees, and trees or shrubs that are small or stunted because of environmental conditions are included.

51 – Shrubland

Areas dominated by shrubs; shrub canopy accounts for 25-100 percent of the cover. Shrub cover is generally greater than 25 percent when tree cover is less than 25 percent. Shrub cover may be less than 25 percent in cases when the cover of other life forms (e.g. herbaceous or tree) is less than 25 percent and shrubs cover exceeds the cover of the other life forms.

Non-Natural Woody - Areas dominated by non-natural woody vegetation; non-natural woody vegetative canopy accounts for 25-100 percent of the cover. The non-natural woody classification is subject to the availability of sufficient ancillary data to differentiate non-natural woody vegetation from natural woody vegetation.

61 - Orchards/Vineyards/Other

Orchards, vineyards, and other areas planted or maintained for the production of fruits, nuts, berries, or ornamentals.

Herbaceous Upland - Upland areas characterized by natural or semi-natural herbaceous vegetation; herbaceous vegetation accounts for 75-100 percent of the cover.

71 - Grasslands/Herbaceous

Areas dominated by upland grasses and forbs. In rare cases, herbaceous cover is less than 25 percent, but exceeds the combined cover of the woody species present. These areas are not subject to intensive management, but they are often utilized for grazing.

Planted/Cultivated - Areas characterized by herbaceous vegetation that has been planted or is intensively managed for the production of food, feed, or fiber; or is maintained in developed settings for specific purposes. Herbaceous vegetation accounts for 75-100 percent of the cover.

81 - Pasture/Hay

Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops.

82 - Row Crops

Areas used for the production of crops, such as corn, soybeans, vegetables, tobacco, and cotton.

83 - Small Grains

Areas used for the production of graminoid crops such as wheat, barley, oats, and rice.

84 - Fallow

Areas used for the production of crops that do not exhibit visible vegetation as a result of being tilled in a management practice that incorporates prescribed alternation between cropping and tillage.

85 - Urban/Recreational Grasses

Vegetation (primarily grasses) planted in developed settings for recreation, erosion control, or aesthetic purposes. Examples include parks, lawns, golf courses, airport grasses, and industrial site grasses.

Wetlands - Areas where the soil or substrate is periodically saturated with or covered with water as defined by Cowardin et al.

91 - Woody Wetlands

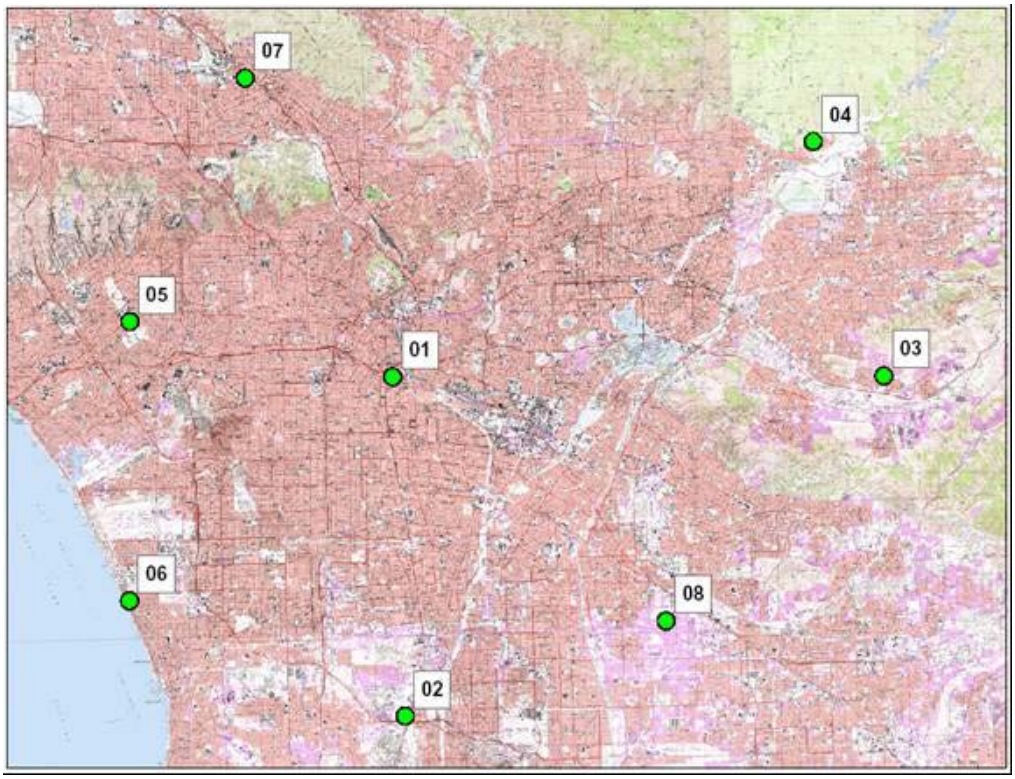
Areas where forest or shrubland vegetation accounts for 25-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.

92 - Emergent Herbaceous Wetlands

Areas where perennial herbaceous vegetation accounts for 75-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.

Investigation of NLCD's Accuracy in Correctly Identifying Farmland

The categories selected from the NLCD dataset to show California farmland were: 61 - Orchards/Vineyards/Other, 82 - Row Crops, 83 - Small Grains, and 84 – Fallow. Grid cells with these values were extracted and added to the California farmland dataset. During an investigation of the resulting composite farmland dataset, agricultural areas were showing up unexpectedly in unlikely places such as downtown Los Angeles and Beverly Hills. In an effort to determine the validity of these unexpected "farmlands," eight study sites and a control site were identified (see map below). These sites were selected in an attempt to provide an array of farmland patterns/groupings as well taking into account the different natural landscapes of an area as large as Los Angeles. The control site, not shown, was located in an area east of investigation site 3.



Methodology

In an effort to keep the analysis simple and limited in scope the following methodology was used:

Step 1 - Identify eight investigation locations and a control site consisting of three separate cluster types within Los Angeles County:

Single isolated parcel
Multiple semi-isolated parcels
Groupings or merged areas of parcels

Step 2 - Identify the NLCD code/description for the selected parcels

Step 3 - Obtain aerial photography from multiple sources and timeframes for comparison:

- Freely downloadable Digital Ortho Quadrangle's provided by the California Spatial Information Library (CaSIL)
- Freely viewable aerial photography provided by MapQuest
- Freely downloadable Landsat 7 satellite data
- Freely downloadable SPOT satellite data

Step 4 - Intensive investigation of acquired data

Use the aerial photography to perform a detailed survey of land use, land features, structures in the area of the study sites. The survey was conducted using a standardized half-mile buffer zone. Graphics and tables were used to support one of 4 possible decisions.

- Obviously Farmland - Identified as Farmland
- Possibly Farmland - On site investigation warranted
- Unlikely Farmland - Possibly an open lot, construction site, etc
- Obvious Error - Water Body, etc

Investigation Sites

Investigation Site 01
Decision: #4 – Obvious Error
Area: Los Angeles
Location: E. Washington Ave / Alameda St.
Description: Single parcel. The NLCD code that is attributed to this site is that of "82" or "Row Crops." This site was chosen due to its status as a single parcel.

Investigation Site 02
Decision: #2 – Possibly Farmland (in 1994)
Area: Dolores
Location: E. Carson St.
Description: Mixed group parcels. The NLCD code that is attributed to this site is that of “82” and “61” or “Row Crops” and “Orchards/Vineyards/Other” This site was chosen due to its status as a group parcel.

Investigation Site 03
Decision: #3 – Unlikely Farmland
Area: West Covina
Location: Stephanie Drive / Stacy Court
Description: Group parcel. The NLCD codes that were attributed to this site was that of “82” or “Row Crops” and “84” or “Fallow.” This site was chosen due to its status as a group parcel.

Investigation Site 04
Decision: #3 – Unlikely Farmland
Area: Duarte
Location: Royal Oaks Dr. / Las Lomas Rd.
Description: Mixed group parcels. The NLCD code that is attributed to this site is that of “82,” “84” and “61” or “Row Crops,” “Fallow” and “Orchards/Vineyards/Other” This site was chosen due to its status as a group of parcels.

Investigation Site 05
Decision: #4 – Obvious Error
Area: Beverly Hills
Location: Beverly Hills City Hall
Description: Individual parcel. The NLCD code that is attributed to this site is that of “82” or “Row Crops.” This site was chosen due to its status as an individual parcel and location.

Investigation Site 06
Decision: #4 - Obvious Error
Area: Manhattan Beach
Location: Highland Ave. / 30th Street
Description: Two individual parcels in a localized area. The NLCD code that was attributed to both was that of “82” or “Row Crops.” This site was chosen due to its proximity to the beach, industrial sites and an airport.

Investigation Site 07
Decision: #3 – Unlikely Farmland
Area: Burbank
Location: W. Empire Ave. / N. Victory Pl.
Description: Scattered Individual parcels. The NLCD code that is attributed to this site is that of "82" or "Row Crops." This site was chosen due to its status as a localized scatter.

Investigation Site 08
Decision: #3 – Unlikely Farmland
Area: Cerritos
Location: 166th Street / Shoemaker Ave.
Description: Group parcel. The NLCD code that is attributed to this site is that of "82" or "Row Crops." This site was chosen due to its status as a group parcel.

Investigation Site – Control Site
Decision: #1 - Obviously Farmland
Area: Diamond Bar
Location: Fountain Springs Rd / S. Diamond Bar Blvd
Description: Large Grouping of NLCD parcels confirmed on 1994 DOQ as Farmland. The NLCD code that was attributed to both was that of “82” or “Row Crops.” This site was chosen due to its confirmed status as farmland in the 1994 DOQ imagery

Conclusion

All cluster types including isolated pixels and even larger contiguous areas appear to be very susceptible to “noise” or classification/reflectance confusion (i.e. classifying lawn areas as agriculture) and are therefore not reliable. Based on the investigation results it was recommended that NLCD not be used as a supplemental farmland data source.

APPENDIX G. FRAR METADATA

Title: Multi-source Land Cover Data (2002 v2)
Catalog Name: California Department of Forestry and Fire Protection
OID: 731

CITATION INFORMATION

Identifier: fveg02_2
Originator: California Department of Forestry and Fire Protection
Publication Date: 2002
Information Resource Type
Format: Computer file
Content: Geographic information system
Scale: See abstract
Other Citation Details: Map projection is Albers Equal Area, NAD27

IDENTIFICATION INFORMATION

Abstract: CDF-FRAP compiled the "best available" land cover into a single data layer, to support the various analyses required for the 2002 Forest and Range Assessment. Typically the most current and detailed data were collected for various regions of the state or for unique mapping efforts (farmland, wetlands, riparian vegetation). Decision rules were developed that controlled which layers were given priority in areas of overlap. Cross-walks were used to compile various sources into the common California Wildlife Habitat Relationships (CWHR) system classification. Data sources had unique scale/resolution, multi source data provided as 100m GRID. The original 1/2002 data used to support the Asessment is also available from the FRAP site - this record corresponds to the most recent updated version of the data (10/2002), which incorporates better data for the Mojave & NE Colorado Desert areas.

Browse Graphic URL: http://frap.cdf.ca.gov/data/browsegraphic/fveg02_1.gif
Purpose: No single source exists that adequately captures the statewide distribution and diversity of land cover. To address the range of issues required in the 2002 Assessment, FRAP compiled best available land cover into a single source.
Supplemental Info: Potential users are encouraged to visit the FRAPVEG site http://frap.cdf.ca.gov/projects/frap_veg/index.html), which has detailed documentation on methodology, links to sites with the source data used in FRAPVEG, update schedule, etc. FRAPVEG statewide and county maps are available for download as well.

Time Period Start:	1/1/1980
Time Period End:	11/1/2002
Currentness:	Ground Condition
Progress:	Complete
Update Frequency:	Irregular
Place:	California
Geographic Region West:	-124
Geographic Region East:	-114
Geographic Region North:	42
Geographic Region South:	32
Themes:	Desert resources, Forests and forestry, Timber, Nontimber forest resources, Biotic communities, Chaparral, Conifers, Deserts, Grasslands, Riparian forests, Vegetation, Woodlands, Land use, Agriculture
User Keywords:	CWHR, Hardwoods, Range Lands, redwood, Shrub, Urban Footprint, Valley Oak
Access Limitations:	No Restrictions
Use Limitations:	FRAP disclaimer Read disclaimer within data dictionary supplied when data are downloaded from FRAP site

Contact Information

Data Contact:	LCMMP Vegetation Mapping Coordinator
Organization:	CDF-FRAP
Phone:	916-227-2658
Fax:	916-227-2672
Email:	Mark.Rosenberg@fire.ca.gov
Url:	http://frap.cdf.ca.gov/
Address:	1920 20th Street 1920 20th Street
City:	Sacramento
State:	California
Postal Code:	95814
Country:	USA

DISTRIBUTION INFORMATION

Online Linkage:	http://frap.cdf.ca.gov/data.html
Size (MB):	60 mb (statewide)
Distribution Format:	ARC/INFO Export
Ordering Instructions:	Obtain on-line or request through distribution contact

Contact Information

Distribution Contact:	FRAP data librarian
Organization:	CDF-FRAP
Phone:	916-227-1381
Fax:	916-227-2672
Url:	http://frap.cdf.ca.gov
Address:	1920 20th Street 1920 20th Street
City:	Sacramento
State:	California
Postal Code:	95814
Country:	USA

METADATA INFORMATION

Date: 10/27/2002

Contact Information

Metadata Contact: LCMMP Vegetation Mapping Coordinator
Organization: CDF-FRAP
Phone: 916-227-2658
Fax: 916-227-2672
Email: Mark.Rosenberg@fire.ca.gov
Url: http://frap.cdf.ca.gov/
Address: 1920 20th Street
1920 20th Street
City: Sacramento
State: California
Postal Code: 95814
Country: USA

Data Dictionary

DATABASE FIELDS:

FIELD NAME	INPUT WIDTH	OUTPUT WIDTH	DATA TYPE	DECIMALS	DESCRIPTION
VALUE	4	10	B	-	
COUNT	4	10	B	-	
WHRNUM	4	16	B	-	unique number for whrtype
WHRNAME	40	35	C	-	long name for whrtype
WHRTYPE	3	3	C	-	wildlife hab relationship code
WHRSIZE	1	1	I	-	whr tree size class code
WHRDENSITY	1	1	C	-	whr tree density class code
WHRDEN_NUM	4	5	B	-	whr tree density number
WHR10NUM	4	5	B	-	major habitat type 10 class number
WHR10NAME	12	12	C	-	major habitat type 10 class name
WHR13NUM	4	5	B	-	major habitat type 13 class number
WHR13NAME	20	20	C	-	major habitat type 13 class name
LIFE_NUM	4	5	B	-	major life form 8 class number
LIFE_FORM	12	12	C	-	major life form 8 class name
SOURCE_NUM	4	16	B	-	source num of vegetation data
SOURCE_NAME	30	30	C	-	source name of vegetation data

FIELD NAME: WHRNUM,WHRNAME, WHRTYPE

California Wildlife Habitat Relationships System (WHR) is "cross walked" from various sources, please review the document "Methods for Development of Habitat Acres Forest and Range Assessment 2002" on-line at http://frap.cdf.ca.gov/projects/frap_veg/index.html

WHRNUM	WHRTYPE	WHRNAME
1	ADS	Alpine-Dwarf Shrub
2	AGR	Irrigated Agriculture
3	AGS	Annual Grassland

4	ASC	Alkali Desert Scrub
5	ASP	Aspen
6	BAR	Barren
7	BBR	Bitterbrush
8	BOP	Blue Oak-Foothill Pine
9	BOW	Blue Oak Woodland
62	CHP	Undetermined Shrub Type
63	CON	Undetermined Conifer Type
10	COW	Coastal Oak Woodland
11	CPC	Closed-Cone Pine-Cypress
12	CRC	Chamise-Redshank Chaparral
13	CSC	Coastal Scrub
14	DFR	Douglas-Fir
15	DRI	Desert Riparian
17	DSC	Desert Scrub
18	DSS	Desert Succulent Shrub
19	DSW	Desert Wash
20	EPN	Eastside Pine
21	EST	Estuarine
77	EUC	Eucalyptus
22	FEW	Freshwater Emergent Wetland
24	JPN	Jeffrey Pine
25	JST	Joshua Tree
26	JUN	Juniper
27	KMC	Klamath Mixed Conifer
28	LAC	Lacustrine
29	LPN	Lodgepole Pine
30	LSG	Low Sage
31	MAR	Marine
32	MCH	Mixed Chaparral
34	MCP	Montane Chaparral
35	MHC	Montane Hardwood-Conifer
36	MHW	Montane Hardwood
37	MRI	Montane Riparian
72	PAS	Non-Irrigated Pasture
39	PGS	Perennial Grassland
40	PJN	Pinyon-Juniper
41	POS	Palm Oasis
42	PPN	Ponderosa Pine
43	RIV	Riverine
44	RDW	Redwood
45	RFR	Red Fir
48	SCN	Subalpine Conifer
49	SEW	Saline Emergent Wetland
50	SGB	Sagebrush
51	SMC	Sierran Mixed Conifer
53	URB	Urban
55	VOW	Valley Oak Woodland
56	VRI	Valley Foothill Riparian
57	WAT	Water
58	WFR	White Fir
59	WTM	Wet Meadow

FIELD NAME: WHRSIZE

California Wildlife Habitat Relationships System (WHR) Tree Size

attribute is "cross walked" from various sources, please review the document "Methods for Development of Habitat Acres Forest and Range Assessment2002" on-line at http://frap.cdf.ca.gov/projects/frap_veg/index.html

TREE type size class code descriptions (Differs from SHRUB size)

WHRSIZE	DESCRIPTION	Diameter at Breast Height (DBH)
0	Not Determined	N/A
1	Seedling	Less Than 1 inch
2	Sapling	1 to 6 inches
3	Pole	6 to 11 inches
4	Small Tree	11 to 24 inches
5	Medium/Large Tree	Greater Than 24 inches
6	Multi Layered	Size 5 Over Size 4 Or 3; Total Tree Crown Closure Greater Than 60%

CWHR Shrub Size Class Descriptions differ from Tree size attribute descriptions. FRAP staff used fire history to model CWHR habitat stages for shrub-dominated habitats. please review the document "Methods for Development of Habitat Acres Forest and Range Assessment 2002" on-line at http://frap.cdf.ca.gov/projects/frap_veg/index.html

SHRUB type size class code descriptions (Differs from Tree size)

WHRSIZE (SHRUB)	Description	Crown Decadence
0	Not Determined	N/A
1	Seedling Shrub	Seedlings or sprouts < 3 years
2	Young Shrub	None
3	Mature Shrub	1 - 25%
4	Decadent Shrub	> 25%

FIELD NAME: WHRDENSITY

California Wildlife Habitat Relationships System (WHR) Density attribute is "cross walked" from various sources, please review the document "Methods for Development of Habitat Acres Forest and Range Assessment 2002" on-line at http://frap.cdf.ca.gov/projects/frap_veg/index.html

WHRDENSITY	WHRDEN_NUM	DESCRIPTION (WHR_RANGE)
	0	None
S	1	10 to 24%
P	2	25 to 39%
M	3	40 to 59%
D	4	60 to 100%

FIELD NAME: WHR10NUM, WHR10NAME

WHR10NUM and WHR10NAME are a hierarchical reclassification of WHR types into 10 "Major Land cover" classes. Statewide maps are available from

our website http://frap.cdf.ca.gov/projects/frap_veg/index.html

WHR10NUM	WHR10NAME
10	Agriculture
20	Barren/Other
30	Conifer
40	Desert
50	Hardwood
60	Herbaceous
70	Shrub
80	Urban
90	Water
100	Wetland

FIELD NAME: WHR13NUM, WHR13NAME

WHR13NUM and WHR13NAME are a hierarchical reclassification of WHR types into 13 "Land cover Subclasses" classes. Statewide maps are available from our website http://frap.cdf.ca.gov/projects/frap_veg/index.html

WHR13NUM	WHR13NAME
10	Agriculture
20	Barren/Other
31	Conifer Forest
32	Conifer Woodland
41	Desert Shrub
42	Desert Woodland
51	Hardwood Forest
52	Hardwood Woodland
60	Herbaceous
70	Shrub
80	Urban
90	Water
100	Wetland

FIELD NAME: LIFE_NUM, LIFE_FORM

LIFE_NUM and LIFE_FORM are a hierarchical reclassification of WHR types into 8 "Life Form Classes" based on morphological differences.

LIFE_NUM	LIFE_FORM
1	CONIFER
2	HARDWOOD
3	SHRUB
4	HERBACEOUS
5	WATER
6	URBAN
7	BARREN/OTHER
8	AGRICULTURE

FIELD NAME: SOURCE_NUM, SOURCE_NAME

FRAP identified the best vegetation/habitat layer for each geographic area and land cover type and assembled them into a statewide habitat layer. Source_num and Source_name allow you to identify the source for each WHRTYPE label. Please review the document "Methods for Development of Habitat Acres Forest and Range Assessment 2002" on-line at http://frap.cdf.ca.gov/projects/frap_veg/index.html

SOURCE_NUM	SOURCE_NAME	Reference Information
1	USGS LAKES	USGS 1:100,000 DLG data modified by Teale Data Center
2	DEVELOP10 - URBAN	1990 Census Block data modified by CDF to reflect uninhabited Public lands
3	DFG-SUISUN MARSH	Department of Fish and Game (DFG) 1999 wetlands data for Suisun Marsh: used to depict habitat stages for this east bay area wetland. http://www.dfg.ca.gov/whdab [This link is no longer active.]
4	CHICO - RIPARIAN	Interested parties please contact the Director of Geographical Information Center, California State University, Chico, Chico, CA, 95929-0425, (530)898-5969 email:cwnelson@csuchico.edu
5	NPS - POINT REYES	National Park Service, Point Reyes National Seashore & GGNRA Vegetation Mapping Program, Point Reyes National Seashore CA National Recreation data was cross-walked to CWHR by CDF and used to depict habitat types for Point Reyes and other areas. Contact Dave Schirokauer, Point Reyes National Seashore, 1 Bear Valley Rd, Point Reyes CA 94950
6	WETLANDS-RIPARIAN	1993 Wetlands and Riparian GIS database, Department of Fish and Game: Selected classes were used to identify wetlands and riparian woody areas in the Sacramento Valley, the San Francisco Bay/Delta, and the San Joaquin Valley. http://maphost.dfg.ca.gov/wetlands/ [This link is no longer active.]
7	FARMLAND MAPPING	Department of Conservation (DOC) Farmland Mapping Program (FMMP) (1998): the classes Prime, State, and Unique were extracted to depict lands currently in agricultural uses. In addition, the Developed class was used as an additional source for locating urban lands. http://www.consrv.ca.gov/dlrp/FMMP/index.html

8	CALVEG 2000	California Department of Forestry and Fire Protection / USDA Forest Service CALVEG (1994 - 1997): used to depict habitats and seral stages on approximately 52 million acres of forest and rangeland areas, including all Region 5 National Forest lands and the Calif. portion of the Toiyabe NF http://frap.cdf.ca.gov/projects/land_cover/mapping/index.html
10	YOSEMITE 1937	Yosemite National Park (1934): Yosemite National Park vegetation data were cross walked to CWHR by CDF and used to represent habitat types for Yosemite NP
11	CDF - HARDWOODS	CDF hardwood rangelands (1990): used for the southern Sierra foothills and central coast areas. http://frap.cdf.ca.gov/data/hardwood_veg/index.html [This link is no longer active.]
12	WETLANDS - GRASS	1993 Wetlands and Riparian GIS database, Department of Fish and Game: Selected classes were used to identify wetlands and riparian woody areas in the Sacramento Valley, the San Francisco Bay/Delta, and the San Joaquin Valley. http://maphost.dfg.ca.gov/wetlands/ [This link is no longer active.]
13	GAP	1998 GAP analysis project, UC Santa Barbara: WHR from this data set was used for areas where no other data was available. Citation: Davis, F. W., D. M. Stoms, A. D. Hollander, K. A. Thomas, P. A. Stine, D. Odion, M. I. Borchert, J. H. Thorne, M. V. Gray, R. E. Walker, K. Warner, and J. Graae. 1998. The California Gap Analysis Project-Final Report. University of California, Santa Barbara, CA.: http://www.biogeog.ucsb.edu/projects/gap/gap_rep.html
14	WETLANDS-ISLANDS	1993 Wetlands and Riparian GIS database, Department of Fish and Game: Selected classes were used to identify wetlands and riparian woody areas in the Sacramento Valley, the San Francisco Bay/Delta, and the San Joaquin Valley. http://maphost.dfg.ca.gov/wetlands/ [This link is no longer active.]
15	MDEP VEG (USGS-BRD)	Central Mojave Vegetation Map: displays vegetation and other land cover types in the eastern Mojave of California. The USGS-BRD

		Developed this as part of the Department of Defense Legacy funded Mojave Desert Ecosystem Program. Data were developed using field visits, and 1997 1:32,000 aerial photography, SPOT satellite imagery, and predictive modeling. The nominal minimum mapping unit is 5 hectares. contact Kathryn Thomas at USGS BRD for more info.
16	NE COLORADO DESERT	(BLM) This data, developed from a combination of GAP Analysis project data, air photos and field visits was generated for the North Eastern Colorado Desert Plan Area in 1997. Contact Nanette Pratini, BLM Riverside at nanette_pratini@ca.blm.gov
17	West Mojave (BLM?)	This data, developed from a combination of GAP Analysis project data, air photos and field visits was generated for the West Mojave Area in (1997?). Contact Nanette Pratini, BLM Riverside at nanette_pratini@ca.blm.gov

COORDINATE SYSTEM DESCRIPTION

Projection	ALBERS		
Datum	NAD27		
Units	METERS	Spheroid	CLARKE1866
Parameters:			
1st standard parallel			34 0 0.000
2nd standard parallel			40 30 0.000
central meridian			-120 0 0.00
latitude of projection's origin			0 0 0.000
false easting (meters)			0.00000
false northing (meters)			-4000000.0000

DISCLAIMER

The State of California and the Department of Forestry and Fire Protection make no representations or warranties regarding the accuracy of data or maps. The user will not seek to hold the State or the Department liable under any circumstances for any damages with respect to any claim by the user or any third party on account of or arising from the use of data or maps.

The user will cite the Department of Forestry and Fire Protection as the original source of the data, but will clearly denote cases where the original data have been updated, modified, or in any way altered from the original condition.

There are no restrictions on distribution of the data by users. However, users are encouraged to refer others

to the Department of Forestry and Fire Protection to acquire the data, in case updated data become available.

DISCLAIMER

The State of California and the Department of Forestry and Fire Protection make no representations or warranties regarding the accuracy of data or maps. The user will not seek to hold the State or the Department liable under any circumstances for any damages with respect to any claim by the user or any third party on account of or arising from the use of data or maps.

The user will cite the Department of Forestry and Fire Protection as the original source of the data, but will clearly denote cases where the original data have been updated, modified, or in any way altered from the original condition.

There are no restrictions on distribution of the data by users. However, users are encouraged to refer others to the Department of Forestry and Fire Protection to acquire the data, in case updated data become available.

APPENDIX H: CALIFORNIA PUBLIC SCHOOLS DATA BASE REPORT

Development of a California Public Schools Data Base with Enhanced Geographic Coordinates

July 27, 2004

Prepared for

U.S. Environmental Protection Agency
Office of Civil Rights (1201A)
1200 Pennsylvania Ave., NW
Washington, DC 20460

Prepared by

ICF Consulting, Inc.
9300 Lee Highway
Fairfax, VA 22031

Contract Number GS-10F-0124J

Disclaimer

Reference in this documentation to any of the following does not necessarily constitute or imply the endorsement, recommendation, or favoring by EPA or the United States Government or their employees or contractors: (1) specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise or (2) other documents, external web sites, or other references and the information contained therein. Furthermore, the views and opinions of authors expressed herein do not necessarily state or reflect those of EPA or the United States Government or their employees or contractors, and shall not be used for advertising or product endorsement purposes. Neither EPA nor the United States Government nor their employees or contractors makes any warranty, expressed or implied, including the warranties of merchantability and fitness for a particular purpose, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights, with respect to documents referenced in this document.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

This document describes a database of California public school information that has been enhanced using GIS-based methods (see box) and other techniques. This database has been developed for use during investigations conducted under the provisions of Title VI of the Civil Rights Act of 1964 for alleged adverse disparate impacts from nearby sources of toxic chemical releases. This database is primarily in the form of a Microsoft Access (version 2003) relational database. An additional form of the data is a GIS shape file, which allows the locations of each school to be available for viewing on a map. The relational database contains either 21 or 22 data tables (the larger version contains a table with enforcement-sensitive information). A number of school years (SYs) are addressed, from SY95-96 through SY01-02, though particular emphasis for enhancement has been placed on school location, demographic, and enrollment data for SY99-00.

Geographical Information Systems (GIS) are computer hardware and software used for storage, retrieval, mapping, and analysis of geographic data. The geographic coordinate data described in this document have been enhanced using a variety of GIS methods.

Below are a description of the background and purpose for this database (Section 2), the approach used to develop the database (Section 3), a summary description of the main database (Section 4), a reference section, and several appendices with more detailed information on the database tables, fields, and code definitions (Attachment 1), the shape file (Attachment 2), and a special Title VI project related to the database (Attachment 3).

For additional information about this documentation or database, contact:

Loren Hall
USEPA Office of Civil Rights (1201A)
Ariel Rios Building
1200 Pennsylvania Avenue, NW
Washington, DC 20460
202-564-0722
hall.loren@epa.gov

2. BACKGROUND AND PURPOSE

The EPA Office of Civil Rights (OCR) is responsible for ensuring compliance with the provisions of Title VI and EPA's implementing regulations. These provisions require that recipients of federal financial assistance operate their programs in a non-discriminatory manner. Persons may file administrative complaints with EPA identifying alleged actions by recipients that result in an adverse disparate impact on the basis of race, color, or national origin. EPA investigates these complaints in order to evaluate these allegations and determine if a violation of Title VI may have occurred. These investigations may involve performing exposure and risk assessments, and, if a finding of violation is made, may also involve identification of pollution prevention and other methods for eliminating or reducing the adverse disparate impact.

EPA's Title VI Draft Revised Investigation Guidance (EPA, 2000) describes several possible steps for such assessments, including determining the scope of an investigation, evaluating the potential for adverse impacts to occur in any affected population, and, if an adverse impact is identified, assessing disparity in impacts and in the composition of affected populations. This document describes a database that could be used in assessing potential impacts to California public school students from nearby sources of toxic chemical releases.

In performing such assessments, receptor location errors can affect the quality of exposure estimates. Because of the cost of obtaining location data in cases where there were rural, erroneous, or missing addresses can be extremely high, it was expected that not all school locations could be identified with equal accuracy. In this effort, schools were organized into three major categories for purposes of setting accuracy goals. The highest-accuracy was needed for schools that were specifically referenced in an investigation, and the associated accuracy goal for these schools was 100 meters or better. The majority of schools were in the middle quality tier, with an accuracy goal of 200 meters. For a small remainder of schools that were not judged likely to be potentially adversely impacted during an investigation, a less stringent minimal accuracy was acceptable, with a goal of 1000 meters.

Many pieces of information about California public schools are needed for analyses of this type. The information needed includes current school location (e.g., address, latitude/longitude), enrollment, and demographic characteristics. However, the vintage and location data in existing databases were found to be inadequate for this purpose. In some databases with high quality location data, some schools known to exist were not included (and schools known to be closed were included), while in others the location data quality was often limited and usually undocumented, etc. Therefore, various enhancements were made to a base set of school data, including additional data collection, review, development, documentation, and formatting, as described below.

3. CALIFORNIA PUBLIC SCHOOL DATA BASE DEVELOPMENT

Two sets of data were identified as key starting points in the development of this inventory of California public schools:

1. California Department of Education (CDE) data, and
2. U.S. Department of Education Common Core of Data (CCD).[39]

The specific CCD database used here had been supplemented previously with address-matched location data by the Agency for Toxic Substances and Disease Registry (ATSDR), using software and data developed by Geographic Data Technology (GDT; see Attachment 1 for

[39] Note that an alternative commonly used definition of this acronym is Census County Division, which is a subcounty geographic area defined by the Census Bureau. This project's use of the Census County Division is described later in this document. For the purposes of this document, however, CCD is used here primarily for Common Core of Data.

more detail, including documentation of the GDT process result codes).[40] Unless otherwise indicated, therefore, the CCD database refers to the data that have been modified by ATSDR. Figures 1 and 2 provide some basic information about each of these two databases.

These databases were compared and evaluated, and a set of records was identified for which enrollment demographic data for SY99-00 were also available. For these schools, improved location and other data were obtained via several data collection methods. The collection method and other key information are listed for each school record in the database.

Figure 1. Characteristics of the CCD Database

- The Common Core of Data (CCD), from which the CCD database (as modified by ATSDR) is derived, is the U.S. Department of Education's primary database on public elementary and secondary education. CCD is a comprehensive, annual, national statistical database of all public elementary and secondary schools and school districts. It contains data that are designed to be comparable across all states.
- The schools identified are from the inventory for school year 1999-2000
- Contains 8,278 school records for California (in version modified by ATSDR).
- Data are originally reported by state education agencies, which are given an opportunity to review and correct errors.
- Updated for each school year (in regular CCD database).
- May exclude preschool-only programs (in version modified by ATSDR).
- Field definitions are located in Attachment 1.

Figure 2. Characteristics of the CDE Database

- California Department of Education (CDE) data include location, demographic, and other descriptive data for California public schools. The location file contains schools that are currently open, schools that have been closed, and schools that have been retired. The demographic files contain data for open schools only (for the given school year).
- Contains 9,393 school records for California (includes adult education and other schools ultimately determined to be outside the scope of enhanced database, but excludes closed, pending, and duplicate schools and school districts) as of 1999. This number may still include some double-counting but it does not include school district records.
- Updated for each school year.
- Includes preschool programs.
- Original field definitions are provided in Attachment 1.

The development of the integrated set of school records, for which improved location and other data were gathered, was conducted using four basic steps:

(1) Establish the initial universe of schools;
(2) Retrieve demographic and enrollment data where available;
(3) Assess and refine the original location information; and
(4) Conduct quality assurance review of school locations.

[40] GDT provides proprietary street data and geocoding services. GDT was acquired in July 2004 by Tele Atlas (i.e., after the use of GDT by ATSDR).

These steps are described in detail below.

3.1. Establish the Initial Universe of Schools

The purpose of this step was to establish an initial list of individual, operating, California public schools for SY99-00. As a preliminary step, the CDE and CCD databases were compared using the unique school identifiers (i.e., the last seven digits of the CDS_CODE field in the CDE database and the seven-digit ST_ID field in the CCD, which are equivalent), with the result that 8,135 school records were found to overlap in the two databases. While a number of the CDE entries clearly represented instances outside the scope of the enhanced database (e.g., school district offices, summer camp programs, etc.), other entries appeared to represent legitimate additional records (compared to the CCD data) of currently open public schools. Similarly, the CCD data appeared to include records for schools not found in the CDE data. Since neither set alone was known to contain either the full set of open schools in SY99-00, records from each were chosen in an effort to create one comprehensive schools database.

The CCD data records were selected as the primary source of identification data (name, address, etc.) because they have more complete location information available compared to the CDE data (see Section 3.3). The CCD database already had been reduced from its original size via modifications by ATSDR, and thus the only records excluded at this point from the original 8,278 were 58 records that were labeled juvenile detention centers, for a result of 8,220 records.

Prior to combining the CCD and CDE records the original CDE records were evaluated against a number of selection criteria in order to exclude records not applicable to the enhanced database. Table 1 describes the specific criteria used to exclude records, and it presents the number of records excluded due to each criterion.

Table 1. Criteria Used to Exclude CDE Records for the Initial Universe of Schools

Criterion	Number of Records Excluded[a]
Original CDE records = 14,386	
School closed, pending, or merged (not open)	3,943
Duplicate records (e.g., school noted as both open and pending)	1,784[b]
School name and ID blank	1,244
School districts (not specific schools)	1,242
Facility is for adult education (not children)	354
Facility is a juvenile detention center (not a school)	213

[a] Excluded records are "double counted" (i.e., a record excluded for more than one criterion would be counted under all such criteria) and thus the total excluded exceed the difference between original and final CDE records.
[b] These duplicate records represent 395 schools.

As described in Section 3.4, the initial California public school list for SY99-00 was checked to confirm the inclusion of "priority" schools previously identified for a specific Title VI

investigation—see Attachment 3, Project Specific Data Documentation (Investigation Sensitive). One additional school was added due to this effort.

Following the demographic data collection and other steps described below in Sections 3.2-3.4, the CCD ultimately provided 23 schools that were not in the CDE, the CDE data ultimately provided 334 records that were not in the CCD, and the priority schools list provided one record that was in neither the CCD nor CDE databases. The total contribution from both these sources for public school locations in California for SY99-00 is 8,492. Attachment 1, and in particular portions of Table 1-1, describes the data tables and fields related to the identities of the California enhanced database schools.

3.2. Collect Demographic and Enrollment Data

Demographic and enrollment data were available from CDE for 8,568 SY99-00 California schools (see http://www.cde.ca.gov/ds/sd/sd/filesethsch.asp). These data were matched with the schools obtained in the previous step, with the result that 75 schools not previously identified were added to the developing database. However, 62 CCD records did not have corresponding demographic information for SY99-00 and thus were excluded for the purposes of the enhanced database.

Also included in the enhanced database is the grade range represented by the school inventory and the number of students in attendance at each school. Additional demographic and enrollment data, for only those schools open in SY99-00 and for which demographic data were available, were obtained for SY95-96 through SY01-02.

The race and ethnicity information obtained from the CDE demographic and enrollment databases are as follows:

- African American (not Hispanic),
- American Indian or Alaskan Native,
- Asian,
- Filipino,
- Pacific Islander,
- Hispanic or Latino,
- White (not Hispanic), and
- Multiple.

This categorization, however, is inconsistent with the 1990 U.S. Census data and OMB Directive 15 (OMB, 1997), which distinguish between race and ethnicity as two separate categories, as follows:

Race

- African American,
- American Indian or Alaskan Native,
- Asian,

- Native Hawaiian or Other Pacific Islander, and
- White.

Ethnicity
- Hispanic or Latino and
- Not Hispanic or Latino.

Consequently, Census and other data allow for combinations of race and ethnicity such as White/Hispanic and African American/Hispanic, while the CDE database does not. Furthermore, the CDE database includes two groups—Pacific Islander and Filipino—where Census and other data use one. Thus, if this database is to be used in conjunction with Census or similar data, then some normalization of the data may be needed.

See Attachment 1, in particular Table 1-1, for data tables and fields related to demographics and enrollment.

3.3. Assess and Refine the Original Location Information

In this step, the location coordinate information (latitude and longitude) for each record in the school universe was assessed and refined as needed using a variety of data sources. Both the CCD and CDE databases have coordinate information, and therefore the first step was to determine which set of coordinates should be used preferentially over the other set. Thus, it was noted that approximately 85% of the coordinates in the CCD database already had been address-matched (i.e., coordinates for specific street addresses had been obtained) by ATSDR using the GDT commercial service (as described above). The remaining coordinates either had been obtained using less accurate approaches, such as the zip code area center point, or in a few cases were blank.

In contrast, it is unknown how many of the records in the CDE database were address-matched because the documentation does not specify location data collection method or accuracy. Furthermore, many of the coordinates in the CDE database either have only two or three decimal places—which would translate into approximately an 800 meter error on the ground—or have no coordinate information at all. Therefore, it was decided that the CCD coordinates should be used preferentially over the coordinates provided with the CDE database.

Information as to which horizontal datum[41] was applied was not provided along with the ATSDR-modified CCD data. GDT documentation indicated that the datum for an address-matching deliverable from GDT is typically the more recent NAD83 datum rather than the older NAD27 datum (otherwise an extra fee would apply for conversion). However, although GDT data and software were used to address-match the CCD school addresses, the actual matching was run internally at ATSDR and not by GDT. Therefore, to determine which datum was

[41] The horizontal datum is the base reference for a coordinate system. It includes the latitude and longitude and orientation of an initial point of origin of an ellipsoid that models the surface of the earth in the region of interest. For example, North American Datum of 1927 (NAD 27) is based on the Clarke ellipsoid of 1866 and a marker in Kansas. NAD83 is an earth-centered datum based on a more recently defined ellipsoid—the Geodetic Reference System of 1980 (GRS80).

actually used for the ATSDR-modified CCD data, a sample of schools were mapped according to different datum assumptions and verified relative to street locations. The schools were first assumed to be in NAD83 and overlain with road information and aerial photos. In a second step, they were assumed to be in NAD27, were converted to NAD83, and then were overlain with road information and aerial photos. The first step resulted in locations that were somewhat distant from the relevant road and school location as compared to the second step, which would indicate that the original datum was NAD27. These results are illustrated in Figure 3 for one of the schools examined. The pentagon (⬠) represents the location under datum NAD27, while the triangle (△) represents the location under datum NAD83. (This photos also shows a "student activity center" (⊙) that was estimated for some schools and the core group of school buildings and facilities ( ) used for the derivation of this center.) This pattern among the sample of schools tested led to the conclusion that the GDT data were provided in NAD27. As a result, all schools from CCD were converted to NAD83.

Figure 3. Example of Different Datum Assumptions

As noted above, for both datasets some coordinates were missing, some were suspect due to decimal truncation, and others were based on zip-code or other centroid geocoding which is too imprecise for the purposes of this study. Therefore, a number of steps were taken at this point to refine the California school location information. These steps involved the use of various address-matching services. Briefly, the key address-matching services used were:

- Tele Atlas, a proprietary address-matching service that provides a useful "batch" approach to obtaining coordinate information. Tele Atlas was used for the largest set of suspect records from CCD and CDE. See www.na.teleatlas.com for more information about Tele Atlas.

- Maporama, which offers a geocoding services with advanced address parsing and address error resolution algorithms. The user is allowed to select from a set of likely candidates

for a match, informed by an interactive vector map interface. See www.maporama.com for more information about Maporama.

- U.S. Geological Survey (USGS) Geographic Names Information System (GNIS), which provides generally higher accuracy location data than address-matched data. See http://geonames.usgs.gov for more information about GNIS.

- Topozone, which allows the input of coordinates against U.S. Geological Survey (USGS) digital raster graphic views. See www.topozone.com for more information about Topozone.

- Terrafly, which provides aerial photo imagery derived from USGS digital orthophoto quadrangles and orthorectified for precise mapping. See www.terrafly.com for more information about Terrafly.

- GlobeXplorer, which provides alternative high resolution aerial photo imagery. See www.globexplorer.com for more information about GlobeXplorer.

These services are discussed below in context of the steps taken to refine the school location data. Table 2 summarizes these steps and provides the numbers of records resulting from each step.

Table 2. Counts of Address-matched Records by Location Source

Source	Number of Resulting Records
Original GDT Coordinates from ATSDR-supplied CCD Database	6,844[a]
Tele Atlas	1,106[a]
GNIS	264[b]
Maporama	143[b]
Terrafly	3[b]
Topozone	3[b]
Activity Center Approximation	8[b]
CDE Coordinates	91
Air Resources Board for California	6
GlobeXplorer	4
Total records = 8,472	

[a] See Table 3a and 3b for breakdowns of these records.
[b] Mainly priority schools, but also some random schools for quality assurance (see Section 3.4).

The combined CCD/CDE inventory of schools initially included 8,479 records. As a first step, the 6,844 records with GDT coordinates from the ATSDR-modified CCD data were retained and set aside. A group of 1,565 records—i.e., schools for which CCD coordinates were not available, schools for which the CCD had zip-generated coordinates, and school records found only in CDE—were sent to a second commercial service (Tele Atlas) for updated address matching to obtain location coordinates where possible. Seven of these records were found to be

duplicates, and thus 1,558 records were returned. Table 3 presents counts of these records organized by address matching code.

Of the 1,124 address-matched records (i.e., with codes 1 or 2 for Block Face and Near Match, respectively) returned from Tele Atlas, 1,106 were used as is while 18 were selected for additional data collection and verification along with a subset of the 498 remaining (poor quality) records. These selected records were chosen based on (1) a random selection of schools for quality control purposes (see Section 3.4) and (2) the overlap between these schools and a set of schools identified as high priority for location verification (see Attachment 3).

The additional location data collection, refinement, and verification for this subset were conducted as follows (see Table 2 for the resulting numbers of records):

1. Use GNIS school location coordinates to replace as many of the coordinates for the high priority schools as possible, as well as for a sample of random non-priority schools that had undergone other address-matching (below). Matches were based on the school names in the given county, since no actual address information is included in GNIS.[42]

2. Use Maporama next to address-match as many of the coordinates as possible. This was conducted in combination with other tools (Topozone, Terrafly) that provide images or maps of the candidate school locations.

3. Use CDE coordinates for the remaining non-priority schools, based on the assumption that the state's coordinates were likely of greater accuracy than the commercially obtained coordinates based only on zip code centroids, although even among this set, an effort was made to prioritize those schools close to farmland and with high school populations so that the coordinates for these might be improved.

4. Derive activity centroids for a small priority group of schools, in lieu of using address-matching services for location refinement. The coordinates were mathematically derived by calculating the central point of the student activity center whose boundaries were digitized from a high resolution aerial photograph. Additionally a few other coordinates were updated from California Air Resources Board information and from GlobeXplorer imagery.

[42] Crosswalking these results with Census TIGER landmark data, which are a subset of GNIS records, was not performed.

Table 3. Address Matching Results

a. GDT Address Matching Results		
GDT Code	**"Matched" Type(s) and Definition**	**Count**
11	*Centroid and blank:* No match: Locality not found in list of valid localities. The Postal Service and therefore GDT does not list this as a deliverable locality.	13
12	*Centroid:* No match: Street address parse error. Incomplete or poorly formatted addresses such as blank fields. The geocoding software was unable to break the address down into prefix, street name, type, suffix, directional, etc. properly.	1
14	*Centroid and blank:* No match: Street name could not be found. The street name given is not found in GDT's street database. This is either due to missing data in GDT's database or an invalid address.	429
15	*Centroid and blank:* No match: Address range did not exist. The address given is not found in GDT's street database on the street given. This is either due to missing address ranges on streets in GDT's database or an invalid address number.	675
16	*Centroid and blank:* More than one segment with address range. Ambiguity: Either due to the address occurring multiple times in GDT's database, or the address isn't specific enough (ex: "100 Main St" when GDT's database contains both "100 N Main St" and "100 S Main St")	47
17	*Centroid:* Unable to match intersection. There may be two valid streets that do not intersect or one or both streets could not be matched. Another possibility is that the two streets intersect in more than one place.	8
18	*Centroid and blank:* Locality not covered by GDT database. The locality is valid, and is listed in GDT's database, but has no streets associated with it. This is typically where the locality and/or ZIP code given corresponds with point delivery. (Ex. Quechee, VT 05059 only refers to a PO Box section in a main post office for the surrounding town of White River Junction, VT 05001.)	78
B1	*Street:* Street segment match	6920
B2	*Street:* Intersection match - based on the "to" end of the first street	20
B3	*Street:* Intersection match (based on the "from" end of the first street)	54
B5	*Street:* Alternate name match: address-match to an alternate or "secondary" name of a street in GDT's database. A common example: when a US or State highway passes through a town, becoming "Main St".	15
B6	*Street:* Placeholder match: matched to a point in GDT's database that had been placed earlier as a "best estimate until acquisition of better resources".	18
Total		8,278

b. Tele Atlas Address Matching Results		
Tele Atlas Code	**Code Type and Definition**	**Count**
CDE Records		
0	*Non-Match:* No match found	1
1	*Block Face:* Match is to within a unique intersection OR match is within a single side of a single street block	258
2	*Near Match:* Match is to a single street block but the correct side of street and correct placement within block are not known	54
3	*ZIP+2 Centroid:* Match is to a ZIP+2 vicinity	2
4	*5-digit ZIP Centroid:* Match is to a 5-digit ZIP vicinity	46
5	*3-digit SCF Centroid:* Match is to a 3-digit ZIP vicinity	3
6	*Ambiguous Match:* Match is to multiple street segments	2
Total		366
CCD Records		
1	*Block Face:* Match is to within a unique intersection OR match is within a single side of a single street block	560
2	*Near Match:* Match is to a single street block but the correct side of street and correct placement within block are not known	252
3	*ZIP+2 Centroid:* Match is to a ZIP+2 vicinity	26
4	*5-digit ZIP Centroid:* Match is to a 5-digit ZIP vicinity	334
5	*3-digit SCF Centroid:* Match is to a 3-digit ZIP vicinity	0
6	*Ambiguous Match:* Match is to multiple street segments	20
Total		1,192
Grand Total		1,558

These data collection methods are central to the level of confidence assigned to each record. Thus, since GDT and Tele Atlas use different codes for similar methods, they had to be standardized for assigning confidence levels and documenting the data collection method. This was done using the crosswalk seen in Table 4 between the GDT and Tele Atlas codes and a standardized address-matching collection method code for this database. Coordinates from GDT-matched records with a match code of poorer than B1 were not used.

Table 4. Crosswalk Between the GDT and Tele Atlas Codes and Standardized Codes

Standard-ized Code	Standardized Description	Tele Atlas Code	Tele Atlas Description	GDT Code	GDT Description	Cross-walk Explanation
A1	Address-match to building number	1	Block Face: Match is to within a unique intersection OR match is within a single side of a single street block	B1	Street segment match	Phone correspondence with Rick Sheridan of GDT Customer Help (800-331-7881), and Tele Atlas Customer Support, confirmed that B1 and 1 are consistent with the definition for A1, despite the seeming incongruity between the official definitions.
A2	Address-match to street block segment	2	Near Match: Match is to a single street block but the correct side of street and correct placement within block are not known	--	--	--

Numerous other location-related data were included in the database. The following set of data fields containing additional location identifiers were populated from the ATSDR-modified CCD database: Tract90; Blkgrp90; Place; MCD; and MSA. In addition, the following location identifier columns were populated using point-in-polygon record update techniques:[43] StateFIPS; CntyFIPS; County; and CCD_Name (for Census County Division, not Common Core of Data). Some of the entries for these fields are blank where the record was sourced from CDE rather than CCD (Common Core of Data). Attachment 1 provides additional detail regarding these location fields, the confidence indicators used for the location data, and the other data fields and tables that are related to location.

3.4. Conduct Quality Assurance Review of School Locations with Poor Coordinates Near Farmland

There were 150 schools for which GDT and Tele Atlas geocoding failed to provide acceptable coordinates (as described in the previous section). These coordinates for these schools were originally drawn from CDE. This source contains undocumented coordinates, which are in many cases truncated to 2 and 3 decimal places, indicating a limit to the precision of the school's relative location. In order to prioritize which school coordinates might be improved from this group, relevance to this project was assessed based on proximity to farmland and size of the school population. Of these 150 schools, [44] 44 were within 6 miles of farmland and had a population greater than 100 students.

The process used to obtain better coordinates for these 44 was as follows:

1. Search for school name and county in GNIS
   a. If found, accept coordinates
   b. If not, continue to next step

2. Search for school via the school's district website (e.g., http://www.mariposa.k12.ca.us/); obtain verified or corrected address

3. Use Maporama internet mapping sites to locate address; extract coordinates from Maporama

4. Verify coordinates with GlobeXplorer digital air photos and develop activity centroids as appropriate

5. Update records with new coordinates and all associated field entries in main database.

[43] This is a topological overlay procedure that determines the spatial coincidence of points within polygon boundaries. Points are assigned the attributes of the polygons within which they fall.

[44] In addition, there was one school that was in the same town as two of the schools that met the criteria but was not included in the subset. This school, Condor Elementary School (SCHOOL_ID = 6109615), was added to the subset.

Using these methods, 40 of the 44 records were updated. Of these, 12 were updated with locations obtained from GNIS and 28 records were updated using address matching. The 4 remaining records were identified as special education and charter schools, which are variously located at changing locations throughout the school district and whose locations therefore could not be updated. See Attachment 1 for a description of the assigned SOURCE_LOC type and the relevant method, accuracy, and description (MAD) code assignments based on EPA standards (EPA, 1994).

4. SUMMARY AND CONCLUSIONS

The database resulting from the above steps includes California public schools active during one school year (SY99-00), together with documented location, and demographic enrollment information. Related tables include information about these schools from other years, including demographic data and status information about the whether the school was open in SY02-03.

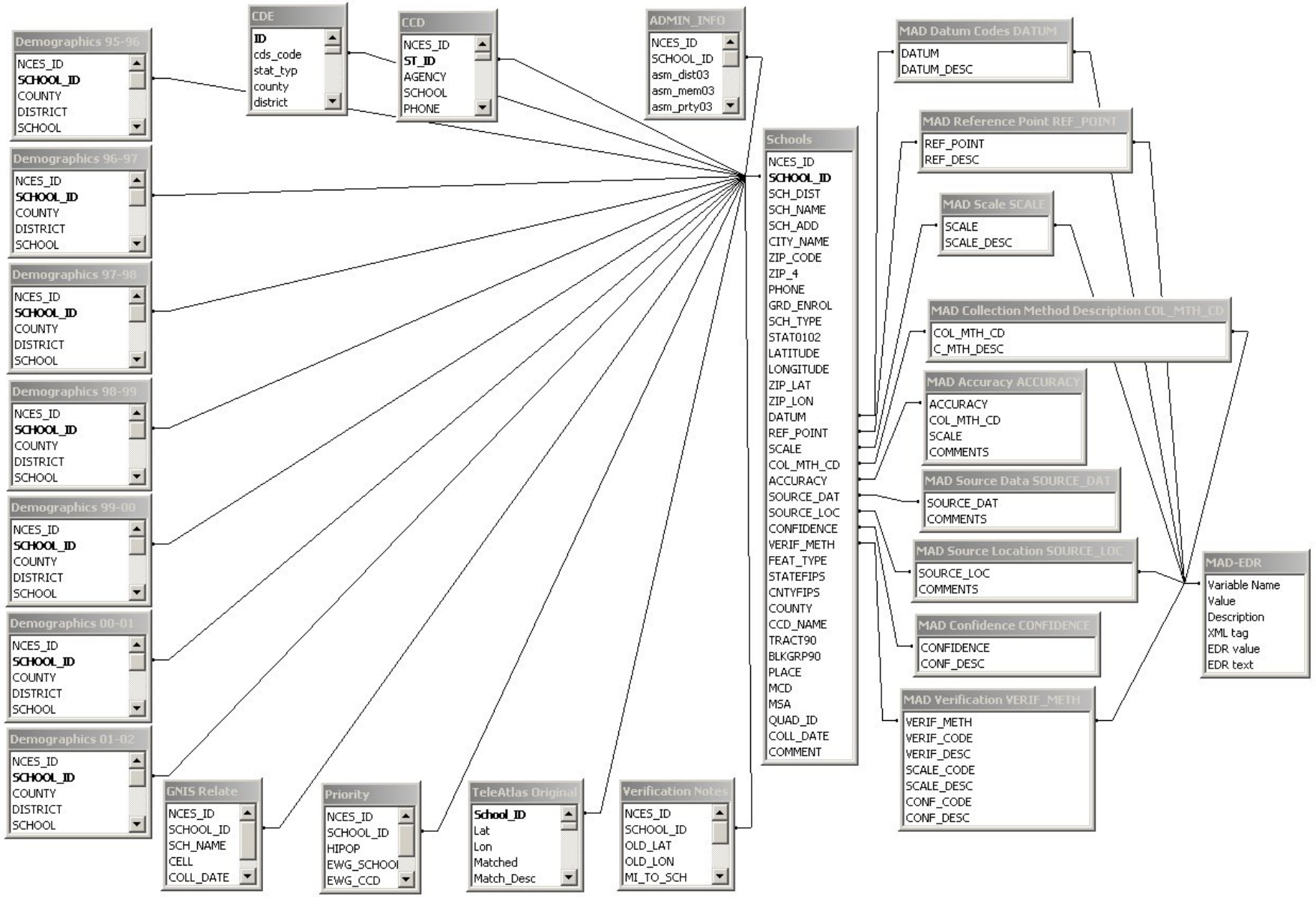
The database structure includes one base table and a number of relate tables, as seen in Figure 4 (the entity relationship diagram, or ERD). The relate tables—primarily details of the location data collection process MAD information—can be used to understand the geocoding method and accuracies associated with the school location information. These tables serve as lookups to data provided in the base schools table.

As described above, an effort was made to capture the most accurate school locations readily available, and to provide documentation concerning the collection method and accuracy for each school. Documentation of data quality within the database includes the coding of each key school record with MAD information based on EPA standards (EPA, 1994). Supporting lookup tables explaining the MAD codes are also included in the database. The MAD tables and codes are described in Attachment 1. Included is reference to a lookup table in the database that crosswalks the standard EPA MAD codes for location information to the codes used in this database (EPA, 1997). See Table 5 for a summary of the breakdown of records by collection method, accuracy, and confidence.

Table 5. Counts of Records By Collection Method, Accuracy, and Confidence

Collection Method	Accuracy (meters)	Confidence			Total Records
		High	Medium	Low	
G3—GPS	10	6	-	-	6
I1—Interpolation from a map	50	264	-	-	264
A1—Address matched to building number	150	22	7,631	-	7,653
A2—Address matched to street block segment	150	145	-	-	145
I2—Interpolation from a photo	25	12	-	-	12
I2—Interpolation from a photo	150	3	-	-	3
A2—Address matched to street block segment	300	1	297	-	298
UN—Unknown	800	-	-	91	91
Total		453	7,928	91	8,472

Figure 4. Entity Relationship Diagram (ERD) for the Geographically Enhanced California Schools Database[a]



[a] Only the MAD data fields are shown in their entirety in this ERD. Furthermore, the CDE and SCHOOLS tables are related by the last seven digits of the CDS_CODE field in the CDE table and the seven-digit SCHOOL_ID field in the SCHOOLS table. See Attachment 1 for additional details about the database tables and fields.

5. REFERENCES

California Department of Education (CDE). No date. File Structure: List of California Public School Districts and Schools. Last accessed on July 27, 2004 at http://www.cde.ca.gov/ds/si/ds/fspubschls.asp.

U.S. Department of Education (ED). 2001. *Documentation to the NCES Common Core of Data, Public Elementary/Secondary School Universe Survey: School Year 1999-2000. (Revised September 2001).* http://nces.ed.gov/ccd/pdf/psu99genr.pdf.

U.S. Environmental Protection Agency (EPA). 1994. *Method Accuracy Description (MAD) (Version 6.1) Information Coding Standards for the U.S. Environmental Protection Agency's Locational Data Policy (LDP).* Prepared by the LDP Sub-Work Group of the Regional GIS Work Group.

U.S. Environmental Protection Agency (EPA). 1997. *Envirofacts Locational Reference Tables, Version 2.1 Data Modeling Report.* [Later replaced with the locational reference tables business data model, at: http://www.epa.gov/enviro/html/locational/lrt/model_business.html]

U.S. Environmental Protection Agency (EPA). 2000. Title VI Draft Revised Investigation Guidance. *Federal Register*. June 27, 2000, Vol. 65, No. 124, pp. 39649-39701.

U.S. Environmental Protection Agency (EPA). 2001. *Standard Data Elements for Latitude/Longitude*. [Later replaced by Environmental Data Standards Council (EDSC) 2006. *Latitude/Longitude Data Standard* available at http://www.exchangenetwork.net/standards/Lat_Long_Standard_08_11_2006_Final.pdf]

Office of Management and Budget (OMB). 1997. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. *Federal Register*. October 30, 1997, Vol. 62, No. 210, pp. 5872-5889.

Sylvan Learning Systems. 2003. http://www.greatschools.net/modperl/go/CA. An online guide to K-12 schools.

GDT. 2002. *Geocoding Users Guide 2002*. [This company was later acquired, and the service was renamed Tele Atlas Geocoding Services, with a more recent manual available at: http://licensing.tomtom.com/stellent/groups/public/documents/content/ta_d_022259.pdf]

Tele Atlas. 2004. Geocoding Documentation. Last accessed on July 27, 2004 at http://www.geocode.com

GNIS. 1995. *Geographic Names Information System, Data Users Guide 6.* [This document has been replaced with links to downloadable data file formats, available at: http://geonames.usgs.gov/domestic/download_data.htm]

ATTACHMENT 1. CALIFORNIA PUBLIC SCHOOLS DATA DICTIONARY

Table Att1-1. Schools Table

Table Name: Schools

Source: ICF Consulting

General Description: This is the main table of California SY99-00 schools to which all other tables relate.

Field Name	Definition
NCES_ID	National Center for Education Statistics (NCES) nationally unique school identification code
SCHOOL_ID	State-wide 7-digit unique school identification code
SCH_DIST	School district name
SCH_NAME	School name
SCH_ADD	School address
CITY_NAME	City name
ZIP_CODE	Zip code
ZIP_4	Zip Plus 4
PHONE	School phone number
GRD_ENROL	Grade span reported to the California Department of Education for this school.
SCH_TYPE	The type of school. As of publication date the possible values were: alternative, Co Community (County Community), community day, continuation, CYA (California Youth Authority), elementary, high school, junior high, juvenile hall, K-12, middle, opportunity, or special Ed. Additional values may include: other county-wide programs, single elementary school in district, single high school in district, state special school, or adult education center.
STAT0102	Status during 2001-2002 school year. As of publication date the possible values were: open, closed, or retired. Additional values may include: merged and pending.
LATITUDE	Latitude coordinate (decimal degrees)
LONGITUDE	Longitude coordinate (decimal degrees)
ZIP_LAT	Latitude of zip code centroid of school as geocoded by Tele Atlas or GDT (decimal degrees)
ZIP_LON	Longitude of zip code centroid of school as geocoded by Tele Atlas or GDT (decimal degrees)
DATUM	Datum code (see Table 1-7, MAD Datum Table)
REF_POINT	Reference point code (see Table 1-8, MAD Reference Point Table)
SCALE	Scale of source for coordinate capture (see Table 1-9, MAD Scale Table)
COL_MTH_CD	Code for how latitude/longitude was obtained (see Table 1-10, MAD Collection Method Description Table)
ACCURACY	Accuracy of latitude/longitude (in meters) (see Table 1-11, MAD Accuracy Table)
SOURCE_DAT	Original source of latitude/longitude and other school attribute data (see Table 1-12, MAD Source Data Table)
SOURCE_LOC	Final source of school latitude/longitude (see Table 1-13, MAD Source Location Table)
CONFIDENCE	Confidence of latitude/longitude (see Table 1-14, MAD Confidence Table)
VERIF_METH	School location verification method (see Table 1-15, MAD Verification Table)
FEAT_TYPE	Geographic feature type. As of publication date the possible value was: point. Additional values may include: line, polygon.
STATEFIPS	Federal Information Processing Standards (FIPS) code for states, assigned by ICF Consulting
CNTYFIPS	County FIPS code, assigned by ICF Consulting
COUNTY	County name, assigned by ICF Consulting

Table Att1-1. Schools Table

Table Name: Schools

Source: ICF Consulting

General Description: This is the main table of California SY99-00 schools to which all other tables relate.

Field Name	Definition
CCD_NAME	Census County Division name (not Common Core of Data). Populated by ICF Consulting using point-in-polygon operation on original school coordinates from CCD (Common Core of Data) and CDE
TRACT90	1990 Census tract from CCD
BLKGRP90	1990 Census block group from CCD
PLACE	Census Place FIPS code for matched point from CCD
MCD	Minor Civil Division from CCD
MSA	Metropolitan Statistical Area from CCD
QUAD_ID	Corresponding U.S. Geological Survey (USGS) 1:24,000 Topographic Map Quadrangle ID used in verifying school location based on Geographic Names Information System (GNIS)
COLL_DATE	Date latitude/longitude was obtained (see Table 1-10, MAD Collection Method Description Table)
COMMENT	General comments about coordinates and verification

Table Att1-2. Demographics Tables

Table Names: Demographics 95-96; Demographics 96-97; Demographics 97-98; Demographics 98-99; Demographics 99-00; Demographics 00-01; and Demographics 01-02

Source: California Department of Education

General Description: These tables provide demographic enrollment information.

Field Name	Definition
NCES_ID	NCES nationally unique school identification code
SCHOOL_ID	State-wide 7-digit unique School identification code
COUNTY	County name
DISTRICT	District name
SCHOOL	School name
AM_IND	Enrollment—American Indian or Alaskan Native: A person having origins in any of the original peoples of North America and who maintains cultural identification through tribal affiliation or community recognition
ASIAN	Enrollment—Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent
PAC_ISLD	Enrollment—Pacific Islander: A person having origins in any of the original peoples of the Polynesian, Micronesian, or Melanesian islands (excludes the Philippine Islands)
FILIPINO	Enrollment—Filipino: A person having origins in any of the original peoples of the Philippine Islands
HISPANIC	Enrollment—Hispanic: A person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish culture or origin - regardless of race
AFR_AM	Enrollment—African American (not Hispanic): A non-Hispanic person having origins in any of the Black racial groups of Africa
WHITE	Enrollment—White (not Hispanic): A non-Hispanic person having origins in any of the original peoples of Europe, North Africa, or the Middle East, e.g. England, Portugal, Egypt, and Iran
TOT_STDNTS	Enrollment—Total number of students

Table Att1-3. GNIS Relate Table

Table Name: GNIS Relate

Source: United States Geological Survey and ICF

General Description: This table provides additional information on the school records whose coordinates were obtained from GNIS. Collection date information for each of these records is the data of the latest revision to the 7.5 minute quadrangle map in which the school is located.

Field Name	Definition
NCES_ID	NCES nationally unique school identification code
SCHOOL_ID	State-wide 7-digit unique School identification code
SCH_NAME	School name
CELL	USGS 7.5 minute quadrangle name
COLL_DATE	Year of most recent revision to quadrangle map
QUAD_ID	USGS 7.5 minute quadrangle ID

Table Att1-4. Tele Atlas Original Table

Table Name: Tele Atlas Original

Source: ICF Consulting

General Description: A subset of schools were chosen to be sent to a commercial service for address matching to obtain location coordinates, if available. This table reports those 1558 school records address-matched using the Tele Atlas geocoding service. The table includes the addresses that were adjusted to a standardized format by the Tele Atlas geocoding process and subsequently copied into the final SCHOOLS database physical address field. This table includes all fields returned from the Tele Atlas address-matching process. For some records, it includes a field called "Fix" that is coded as "Y" where an operator edited apparent errors in the school address (e.g. to remove city names from the address field) to improve the address match results.

Field Name	Definition
School_ID	State-wide 7-digit unique school identification code
Lat	Original latitude from source database
Lon	Original longitude from source database
Matched	Matched field from the Agency for Toxic Substances and Disease Registry (ATSDR) address matching of CCD database
Match_Desc	Match description of CCD and California Department of Education (CDE) match methods added by ICF Consulting
Street_inp	Original street address from source database
City_inp	Original city from source database
State_inp	Original state from source database
Zip_inp	Original zip code from source database
Match_coun	Tele Atlas match county field
Match_type	Tele Atlas match type. Possible values: Code Type Description 0 Non-Match No match found. 1 Block Face Match is to within a unique intersection OR match is within a single side of a single street block 2 Near Match Match is to a single street block but the correct side of street and correct placement within block are not known 3 ZIP+2 Centroid Match is to a ZIP+2 vicinity 4 5-digit ZIP Centroid Match is to a 5-digit ZIP vicinity

Table Att1-4. Tele Atlas Original Table

Table Name: Tele Atlas Original

Source: ICF Consulting

General Description: A subset of schools were chosen to be sent to a commercial service for address matching to obtain location coordinates, if available. This table reports those 1558 school records address-matched using the Tele Atlas geocoding service. The table includes the addresses that were adjusted to a standardized format by the Tele Atlas geocoding process and subsequently copied into the final SCHOOLS database physical address field. This table includes all fields returned from the Tele Atlas address-matching process. For some records, it includes a field called "Fix" that is coded as "Y" where an operator edited apparent errors in the school address (e.g. to remove city names from the address field) to improve the address match results.

Field Name	Definition
	5 3-digit SCF Centroid Match is to a 3-digit ZIP vicinity 6 Ambiguous Match Match is to multiple street segments
Match_db	Tele Atlas Match_db field
Std_addr	Tele Atlas standardized street address
Std_city	Tele Atlas standardized city
Std_state	Tele Atlas standardized state
Std_zip	Tele Atlas standardized zip code
Std_p4	Tele Atlas standardized zip plus 4
Std_dpbc	Tele Atlas Std_dpbc field
Std_carri	Tele Atlas Std_carri field
Match_addr	Tele Atlas matched street address
Match_city	Tele Atlas matched city
Match_stat	Tele Atlas matched state
Match_zip	Tele Atlas matched zip code
Match_lat	Tele Atlas matched latitude
Match_lon	Tele Atlas matched longitude
Fix	Address from original source database was fixed by Tele Atlas operator from EPA Region 9 office where populated with “Y” in order to improve hit rate.

Table Att1-5. Verification Notes Table

Table Name: Verification Notes

Source: ICF Consulting

General Description: This table records the results from a verification process that compared a small number of school locations with digital USGS quadrangle map images ("digital raster graphic" or DRG), using the online Topozone web site. The fields record the relative distance and direction between the coordinates originally in the SCHOOLS database, and locations of school feature symbols (ideally identified as the same school in annotation) on the map. (http://www.topozone.com).

Field Name	Definition
NCES_ID	NCES nationally unique school identification code
School_ID	State-wide 7-digit unique School identification code
OLD LAT	Latitude from original source data
OLD LON	Longitude from original source data
MI_TO_SCH	Measured distance between coordinates for school in database and coordinates for school in verification source (in miles)
DIRECTION	Approximate compass direction between coordinates for school in database and coordinates for school in verification source
NEW LAT	New latitude based on verification source
NEW LON	New longitude based on verification source
SOURCE	Verification source for new coordinates
COMMENT	General comments about verification

Table Att1-6. Administration Information Table

Table Name: Admin_Info

Source: All fields from California Department of Education schools 1999-2000 database except for unique 7-digit School_ID and NCES-ID added later by ICF Consulting for purposes of easy lookup.

General Description: This relate table captures administrative boundary information available in the CDE data source for many records in the school universe.

Field Name	Definition
NCES-ID	NCES nationally unique school identification code
SCHOOL_ID	State-wide 7-digit unique School identification code
Asm_dist03	Number of the California Assembly district in which this school is located
Asm_mem03	Name of the California Assembly member for the district in which this school is located
Asm_prty03	Political party affiliation of this Assembly member
Sen_dist03	Number of the California Senate district in which this school is located
Sen_mem03	Name of the California Senate member for the district in which this school is located
Sen_prty03	Political party affiliation of this Senate member
Boe_dist03	Number of the California Board of Education district in which this school is located
Boe_mem03	Name of the California Board of Education member for the district in which this school is located
Boe_prty03	Political party affiliation of this Board of Education member
Con_dist03	Number of the U.S. congressional district in which this school is located
Con_mem03	Name of the U.S. congressional member for the district in which this school is located
Con_prty03	Political party affiliation of this congressional member

Table Att1-7. MAD Datum Codes Table

Table Name: MAD Datum Codes DATUM

Source: ICF Consulting

General Description: This relate table contains prospective datum codes and their descriptions. As with other MAD fields and prospective values, the application of the datum codes was informed by the latest EPA standards (EPA, 1994).

Field Name	Definition
DATUM	Horizontal datum of latitude/longitude points. Possible values at publication date were as follows: Value — Datum_Desc 83 — NAD83 (Additional datum values—not used here—include NAD27.)
DATUM_DESC	Datum description (see above)

.

Table Att1-8. MAD Reference Point Table

Table Name: MAD Reference Point REF_POINT

Source: ICF Consulting

General Description: This relate table contains prospective reference point codes and descriptions. As with other MAD fields and prospective values, the application of the reference point codes was informed by the latest EPA standards (EPA, 1994). For example, block face street address-matched coordinates are coded with "GE" for general entrance

Field Name	Definition
REF_POINT	Reference point at which coordinates are collected. Possible values: Value — Description (Ref_Desc) AB — Administrative building CP — Approximate center of student activities GE — General entrance UN — Unknown
REF_DESC	Description of reference point (see above)

.

Table Att1-9. MAD Scale Table

Table Name: MAD Scale SCALE

Source: ICF Consulting

General Description: This relate table contains prospective scale codes and their descriptions. As with other method, accuracy and description fields and prospective values, the application of the scale codes was informed by the latest EPA standards (EPA, 1994). For example, 1:100,000 is the most common scale associated with street databases used to support address-matching procedures

Field Name	Definition
SCALE	Map scale of latitude/longitude source. Possible values: Value — Description (Scale_Desc) 100,000 — 1:100,000 24,000 — 1:24,000 100 — 1:100 NA — Scale not applicable U — Unknown
SCALE_DESC	Description of scale (see above)

Table Att1-10. MAD Collection Method Description Table

Table Name: MAD Collection Method Description COL_MTH_CD

Source: ICF Consulting

General Description: This relate table contains prospective collection method codes and their descriptions, as applicable to the schools data. As with other method, accuracy and description fields and prospective values, the application of the collection method codes was informed by the latest EPA standards (EPA, 1994).

Field Name	Definition
COL_MTH_CD	Code for how Latitude/Longitude was obtained. Possible values: Value — Description (C_Mth_Desc) A1 — Address matched to building number A2 — Address matched to street block segment G3 — Global Positioning System-Wide Area Augmentation System (GPS_WAAS) I1 — Interpolation from a Map I2 — Interpolation from a Photo UN — Unknown
C_MTH_DESC	Description of collection method (see above)

Table Att1-11. MAD Accuracy Table

Table Name: MAD Accuracy ACCURACY

Source: ICF Consulting

General Description: This relate table contains prospective accuracy values (meters) with their associated collection method codes and scales. As with other method, accuracy and description fields and prospective values, the application of the accuracy codes was informed by the latest EPA standards (EPA, 1994).

Field Name	Definition
ACCURACY	Accuracy of latitude/longitude (in meters). Accuracy is a function of the collection method and the map scale. Possible values: Value — COL_MTH_CD — Scale — Comments 10 — G3 — NA 150 — A1 — 100,000 150 — I2 — 100,000 25 — I2 — 12,000 150 — A2 — 100,000 300 — A2 — 100,000 50 — I1 — 24,000 800 — UN — 100,000 — Accuracy estimate based on unknown source coordinates truncated to 2 or 3 decimal places
COL_MTH_CD	How latitude/longitude were obtained (see above)
SCALE	Map scale (see above)
COMMENTS	General comments about accuracy (see above)

Table Att1-12. MAD Source Data Table

Table Name: MAD Source Data SOURCE_DAT
Source: ICF Consulting
General Description: This relate table reports the acronym codes and complete names for data sources used in the creation of the school universe. The only two sources used here were the CCD and CDE databases.

Field Name	Definition	
SOURCE_DAT	Original data source of latitude/longitude coordinates. Possible values:	
	Value	Comments
	CCD	Department of Education Common Core of Data
	CDE	California Department of Education
	GREATSCHOOLS.NET	http://www.greatschools.net
	SCHOOL	Phone call to school
	SCHOOL DIST	School District's website
Comments	Description of source (see above)	

Table Att1-13. MAD Source Location Table

Table Name: MAD Source Location SOURCE_LOC
Source: ICF Consulting
General Description: This relate table reports the prospective sources for school location information. While the source data field reports the source of the non-coordinate attribute data in any given school record, the source location reports the source specifically used to obtain the coordinate information.

Field Name	Definition	
SOURCE_LOC	Final source or derivation of latitude/longitude coordinates. Possible values:	
	Value	Comments
	CA_CDE	California Department of Education
	CA_ARB	California Air Resources Board
	CN_GDT	ATSDR GDT address-matched version of CCD data
	CN_GLOBEXPLORER	GlobeXplorer
	CN_MAPORAMA	Maporama
	CN_ACTIVITY_CENTROID	Approximate center of student activity
	CN_TeleAtlas	Tele Atlas
	CN_TERRAFLY	Terrafly
	CN_TOPOZONE	Topozone
	USGS_GNIS	United States Geological Survey (USGS) Geographic Names Information System (GNIS)
Comments	Description of source (see above)	

Table Att1-14. MAD Confidence Table

Table Name: MAD Confidence CONFIDENCE

Source: ICF Consulting

General Description: This relate table contains prospective confidence codes and their descriptions. As with other method, accuracy and description fields and prospective values, the application of the confidence codes was informed by the latest EPA standards (EPA, 1994).

Field Name	Definition
CONFIDENCE	Confidence of Latitude/Longitude. Possible values: Value — Description (Conf_Desc) H — High—photo verified geocoding, photo interpreted school locations, student activity centers, and GNIS records M — Medium—for geocoded-only records (not photo verified) L — Low—for records not assigned H or M; used primarily for the CDE records because of the uncertainty of the truncated coordinates
CONF_DESC	Description of confidence code (see above)

Table Att1-15. MAD Verification Table

Table Name: MAD Verification VERIF_METH

Source: ICF Consulting

General Description: This relate table contains prospective verification codes and their descriptions. As with other method, accuracy and description fields and prospective values, the application of the verification codes was informed by the latest EPA standards (EPA, 1994). This field is only populated for a subset of schools whose coordinates were verified using alternative methods.

Field Name	Definition
VERIF_METH	Verification hybrid code. Possible values (Verification code+Scale_code+Confidence code): Value — Verif Code — Verif_Desc — Scale Code — Scale_Desc — Conf Code — Conf_Desc F100H — F — Verified relative to map features (Photo) — 100 — 1:100 — H — High Confidence KNAH — K — Ground Truth Conducted (GPS) — NA — Not Applicable — H — High Confidence FUH — F — Verified relative to map features (Photo) — U — Unknown — H — High Confidence FUM — F — Verified relative to map features (Photo) — U — Unknown — M — Medium Confidence R100H — R — Verified relative to map features (Roads) — 100,000 — 1:100,000 — H — High Confidence R100M — R — Verified relative to map features (Roads) — 100,000 — 1:100,000 — M — Medium Confidence R24H — R — Verified relative to map features (Roads) — 24,000 — 1:24,000 — H — High Confidence R24M — R — Verified relative to map features (Roads) — 24,000 — 1:24,000 — M — Medium Confidence
VERIF_CODE	Code detailing how school location was verified (see above)
VERIF_DESC	Description of how point was verified (see above)
SCALE_CODE	Map scale code (see above)
SCALE_DESC	Description of map scale code (see above)
CONF_CODE	Confidence of latitude/longitude (see above)
CONF_DESC	Description of confidence (see above)

Table Att1-16. CCD Table

Table Name: CCD

Source: Department of Education, 1999-2000 Common Core of Data (CCD) for Public Schools, as modified by ATSDR

General Description: This table constitutes the California records from the most complete geocoded national set of schools sourced from the Department of Education with no end user restrictions and with data clean-up and address-matching performed by ATSDR.

Field Name	Definition
NCES_ID	NCES nationally unique school identification code
ST_ID	State-wide 7-digit unique School identification code
AGENCY	Name of the education agency which operates this school
SCHOOL	Name of the school
PHONE	Telephone number of school
MAILADD	The mailing address of the school. It may be a street address or a Post Office box number
MAILCITY	City name of the mailing address
MAILSTAT	Two-letter U.S. Postal Service abbreviation of the state where the mailing address is located
MAILZIP	Five-digit U.S. Postal Service ZIP code for the mailing address
PHYSICAL	Location Address
PCITY	Location City
PSTATE	Location State (PO Abbreviation)
PZIP	Location 5 digit ZIP Code
SCHTYP	NCES code for type of school: Code — Description 1 — Regular school 2 — Special education school 3 — Vocational school 4 — Other/alternative school
OP	NCES code for the school operational status: Code — Description 1 — School was operational at the time of the last report and is currently operational 2 — School has closed since the time of the last report 3 — School has been opened since the time of the last report 4 — School was operational at the time of the last report but was not on the CCD list at that time 5 — School was listed on previous year's CCD school universe as being affiliated with a different education agency
GRADE	School grades offered. First two positions indicate the lowest grade. Second two positions indicate the high grade. The following codes are used for each two positions: Code — Description UG — Ungraded PK — Prekindergarten KG — Kindergarten 01 - 12 — 1st through 12th grade 00 — School had no students reported
STREETADD	Street address used for ATSDR's location match
CITY	City used for ATSDR's location match
STATE	State used for ATSDR's location match
ZIP	Five digit zip code used for ATSDR's location match
ZIP4	Four digit zip code suffix used for ATSDR's location match
LATITUDE	Matched latitude
LONGTUDE	Matched longitude

Table Att1-16. CCD Table

Table Name: CCD

Source: Department of Education, 1999-2000 Common Core of Data (CCD) for Public Schools, as modified by ATSDR

General Description: This table constitutes the California records from the most complete geocoded national set of schools sourced from the Department of Education with no end user restrictions and with data clean-up and address-matching performed by ATSDR.

Field Name	Definition
STATEFIP	State FIPS code for matched point
CNTYFIPS	County FIPS code for matched point
TRACT90	Census Tract FIPS code for matched point
BLKGRP90	Census Block Group FIPS code for matched point
PLACE	Census Place FIPS code for matched point
MCD	Minor Civil Division
MSA	Census Metropolitan Statistical Area FIPS code for matched point
MATCHED	Designation of street address or zip code centroid matched point: Code — Description Street — Street address matched point Centroid — Zip code centroid matched point
STATUS	Designation of matched point specificity: Code — Description B1 — Street segment match. B2 — Intersection match (based on the "to" end of the first street). B3 — Intersection match (based on the "from" end of the first street). B5 — Alternate name match: address-match to an alternate or "secondary" name of a street in GDT's database. A common example: when a US or State highway passes through a town, becoming "Main St". B6 — Placeholder match: matched to a point in GDT's database that had been placed earlier as a "best estimate until acquisition of better resources". 11 — No match: Locality not found in list of valid localities. The Postal Service and therefore GDT does not list this as a deliverable locality. 12 — No match: Street address parse error. Incomplete or poorly formatted addresses such as blank fields. The geocoding software was unable to break the address down into prefix, street name, type, suffix, directional, etc. properly. 14 — No match: Street name could not be found. The street name given is not found in GDT's street database. This is either due to missing data in GDT's database or an invalid address. 15 — No match: Address range did not exist. The address given is not found in GDT's street database on the street given. This is either due to missing address ranges on streets in GDT's database or an invalid address number. 16 — More than one segment with address range. Ambiguity: Either due to the address occurring multiple times in GDT's database, or the address isn't specific enough (ex: "100 Main St" when GDT's database contains both "100 N Main St" and "100 S Main St"). 17 — Unable to match intersection. There may be two valid streets that do not intersect or one or both streets could not be matched. Another possibility is that the two streets intersect in more than one place. 18 — Locality not covered by GDT database. The locality is valid, and is listed in GDT's database, but has no streets associated with it. This is typically where the locality and/or ZIP code given corresponds with point delivery. (Ex. Quechee, VT 05059 only refers to a PO Box section in a main post office for the surrounding town of White River Junction, VT 05001.).

Table Att1-17. CDE Table

Table Name: CDE

Source: California Department of Education

General Description: This file is a list of California public schools. The contact phone number for this data set is (916) 327-0214.

Field Name	Definition
ID	Unique ID added in Access
CDS_CODE	This 14-digit code is the official, unique identification of a school within California. The first two digits identify the county, the next five digits identify the school district, and the last seven digits identify the school. Please note that a CDS code ending in '0000000' indicates a district record not a school.
STAT_TYP	This field identifies the status of the school: open, closed, retired, pending
COUNTY	County name
DISTRICT	District name. Also: • The word UNION in the name of an elementary school district indicates that it was formed from two or more districts. • The word JOINT in a district's name indicates that it includes territory from more than one county.
SCHOOL	School name
STR_ADDR	The street address (physical address) of this school. Caution: Some schools (primarily closed or retired schools) may not have data in this field.
STR_CITY	The city for the street address of this school. Caution: Some schools (primarily closed or retired schools) may not have data in this field.
STR_ZIP	The zip code for the street address of this school. Caution: Some schools (primarily closed or retired schools) may not have data in this field.
MAIL_ADR	The mailing address of this school (i.e., P.O. Box). Caution: Many schools have not provided a mailing address that is distinct from street address. If you are using this file to create mailing labels, we advise that you replace missing data in this field with data from the STR_ADDR field. For your convenience we have filled the previously unpopulated MAIL_ADDR cells with STR_ADDR data.
MAIL_CTY	The city for the mailing address of this school. Caution: Many schools have not provided the city for the mailing address that is distinct from the city of the street address. If you are using this file to create mailing labels, we advise that you replace missing data in this field with data from the STR_CITY field. For your convenience we have filled the previously unpopulated MAIL_CITY cells with STR_CITY data.
MAIL_ZIP	The zip code for the mailing address of this school. Caution: Many schools have not provided the zip code for the mailing address that is distinct from the zip code of the street address. For your convenience we have filled the previously unpopulated MAIL_ZIP cells with STR_ZIP data.
STATE	The state name for all schools is 'CA' for California. Exception: One school in California maintains an Oregon mailing address. For this school the state name is 'OR'.
DST_TYPE	The type of school district. See below for a brief description of the different district types: • A UNIFIED school district includes both elementary and high school educational levels. • An ELEMENTARY school district usually includes kindergarten and grades one through six or eight. • A HIGH school district usually includes grade nine and above but may include grade seven and above.
SCH_TYPE	The type of school.. Types of public schools include Special Education School, County Community School, Youth Authority Facility, Opportunity School, Juvenile Court School, Other County-Wide Programs, Elementary School, Single Elementary School in District, Intermediate/Middle School, Alternative School, Junior High School, K-12 School, High School, Single High School in District, Continuation High School, Community Day School, State Special School, and Adult Education Center.
CHARTER	This field identifies charter schools. The field is coded as follows:

Table Att1-17. CDE Table

Table Name: CDE

Source: California Department of Education

General Description: This file is a list of California public schools. The contact phone number for this data set is (916) 327-0214.

Field Name	Definition
	Code — Description Y — The school is a charter, but not a State Board of Education sponsored charter school. S — The school is a State Board of Education sponsored charter school. N — The school is not a charter.
ASM_DIST	Number of the California Assembly district in which this school is located
ASM_MEM	Name of the California Assembly member for the district in which this school is located
ASM_PRTY	Political party affiliation of this Assembly member: G = Green; R = Republican; D = Democrat
SEN_DIST	Number of the California Senate district in which this school is located
SEN_MEM	Name of the California Senate member for the district in which this school is located
SEN_PRTY	Political party affiliation of this Senate member: R = Republican; D = Democrat
BOE_DIST	Number of the California Board of Equalization district in which this school is located
BOE_MEM	Name of the California Board of Equalization member for the district in which this school is located
BOE_PRTY	Political party affiliation of this Board of Equalization member: R = Republican; D = Democrat
CON_DIST	Number of the U. S. congressional district in which this school is located
CON_MEM	Name of the U. S. congressional member for the district in which this school is located
CON_PRTY	Political party affiliation of this U. S. congressional member: R = Republican; D = Democrat
LATITUDE	The angular distance (expressed in degrees) between the location of this school and the equator measured north to south
LONGITUD	The angular distance (expressed in degrees) between the location of this school and the prime meridian (Greenwich, England) measured from west to east.
CTRACT90	U. S. Census tract number, 1990. The tract number is used to uniquely identify a census tract within a county.
CBLOCK90	U. S. Census block group, 1990. A block is a geographic area bounded on all sides by visible or non visible features shown on U. S. census maps. A block is the smallest geographic entity for which the U. S. Census Bureau collects and tabulates decennial census information.
POP_STAT	This field classifies the location of a school relative to seven categories of populous areas. The categories, descriptions, and codes are listed below. The data in this field are provided by the U.S. Census Bureau. It may take 1-2 years to get a designation for a new school. Code — Classification — Description 1 — Large City — A central city of Consolidated Metropolitan Statistical Area (CMSA) with the city having a population greater than or equal to 250,000. 2 — Mid-size City — A central city of a CMSA or Metropolitan Statistical Area (MSA), with the city having a population less than 250,000. 3 — Urban Fringes of Large City — Any incorporated place, Census Designated Place, or non-place territory within a CMSA or MSA of a Large City and defined as urban by the Census Bureau. 4 — Urban Fringes of Mid-size City — Any incorporated place, Census Designated Place, or non-place territory within a CMSA or MSA of a Mid-size City and defined as urban by the Census Bureau. 5 — Large Town — An incorporated place or Census Designated Place with a population greater than or equal to 25,000 and located outside a CMSA or MSA. 6 — Small Town — An incorporated place or Census Designated Place with a population less than 25,000 and greater than 2,500 and located outside a CMSA or MSA. 7 — Rural, outside — Any incorporated place, Census Designated Place, or non-place

Table Att1-17. CDE Table

Table Name: CDE

Source: California Department of Education

General Description: This file is a list of California public schools. The contact phone number for this data set is (916) 327-0214.

Field Name	Definition		
		MSA	territory designated as rural by the Census Bureau.
	8	Rural, inside MSA	Any incorporated place, Census Designated Place, or non-place territory within a CMSA or MSA of a Large or Mid-Size City and defined as rural by the Census Bureau.
	[blank]	Data not available	New school not yet assigned a population status code by the Census Bureau, or where there has been a change in the CDS Code, or a school not reporting on the California Basic Educational Data System (CBEDS) collection.
CSIS_CON	The California School Information System (CSIS) Consortium to which the district belongs. CSIS is a program to facilitate the transfer of information within the K-12 public school system. For more information about CSIS please visit their web site at www.csis.k12.ca.us/. Each CSIS consortium is composed of districts using the same student information system, so these groupings of districts may not be regional.		
CSISAGNT	The 7-digit county-district code for the district or county office that serves as the fiscal agent for the CSIS consortium to which this district belongs.		
CSIS_PH	The CSIS phase.		
CSIS_WV	The CSIS wave.		
GRD_SPAN	The lowest grade and the highest grade in which student enrollment was reported in the most recent certified CBEDS data. This field may not represent the actual range of grades that a school supports. Note: Special programs at schools such as Independent Study, Alternative Education, and Special Education will often expand the grade span beyond that which is typically considered the grade span for schools of that type (i.e. high schools may have a grade span of 3-12).		

Table Att1-18. MAD-EDR Table

Table Name: MAD-EDR

Source: ICF Consulting

General Description: This table serves as a cross-reference between the EPA standard method accuracy description values and related codes and the codes and descriptions used in this project. For ease of use, the EPA XML tags have been provided as well

Field Name	Definition
Variable Name	MAD Code
Value	MAD Code value
Description	Description of MAD Code value
XML tag	XML tag
EDR value	Environmental Data Registry value
EDR text	Environmental Data Registry value meaning

ATTACHMENT 2. DESCRIPTION OF SCHOOLS SHAPE FILE

Identification_Information:
  Citation:
    Citation_Information:
      Originator: Loren Hall, EPA Office of Civil Rights
      Publication_Date: July 2004
      Title: Schools
      Geospatial_Data_Presentation_Form: vector digital data
      Online_Linkage: Description:
    Abstract: This spatial database contains point locations for public California schools for the 1999-2000 school year. The schools represented in this data set were collected from two primary data sources. The first was the US Department of Education's Common Core of Data (CCD) for public schools that had been obtained and partially assigned locations by the Department of Health and Human Services Agency for Toxic Substances and Disease Registry (ATSDR). The second was the California Department of Education (CDE) school directory database. Retrievals from these databases were made to represent the 1999-2000 school year. The ATSDR extract of the CCD data was thought to have excluded schools that included only Pre-K grades, and possibly other records, so the list was supplemented to include additional schools from the CDE directory, except for those that were identified as closed or retired. Supplemental location data were collected using a combination of techniques. This process included adding method, accuracy and collection reference point description codes describing the location data to each school record, and in most cases, sets of other school location attributes, such as county, metropolitan area, Census tract and block group, and congressional district. Additional data that can be obtained and that are contained in relate tables include California state school-specific demographic information for multiple years.
    Purpose: The EPA Office of Civil Rights (OCR) is responsible for ensuring compliance with the provisions of Title VI and EPA's implementing regulations. These provisions require that recipients of federal financial assistance operate their programs in a non-discriminatory manner. Persons may file administrative complaints with EPA identifying alleged actions by recipients that result in an adverse disparate impact on the basis of race, color, or national origin. EPA investigates these complaints in order to evaluate these allegations and determine if a violation of Title VI may have occurred. These investigations may involve performing exposure and risk assessments, and, if a finding of violation is made, may also involve identification of pollution prevention and other methods for eliminating or reducing the adverse disparate impact.

EPA's Title VI Draft Revised Investigation Guidance describes several possible steps for such assessments, including determining the scope of an investigation, evaluating the potential for adverse impacts to occur in any affected population, and, if an adverse impact is identified, assessing disparity in impacts and in the composition of affected populations. This document describes a database that could be used in assessing potential impacts to California public school students from nearby sources of toxic chemical releases.

In performing such assessments, receptor location errors can affect the quality of exposure estimates. Because of the cost of obtaining location data in cases where there were rural, erroneous or missing addresses can be

extremely high, it was expected that not all school locations could be identified with equal accuracy. In this effort, schools were organized into three major categories for purposes of setting accuracy goals. The highest-accuracy was needed for schools that were specifically referenced in an investigation, and the associated accuracy goal for these schools was 100 meters or better. The majority of schools were in the middle quality tier, with an accuracy goal of 200 meters. For a small remainder of schools that were not judged likely to be potentially adversely impacted during an investigation, a less stringent minimal accuracy was acceptable, with a goal of 1000 meters.

Many pieces of information about California public schools are needed for analyses of this type. The information needed includes contemporary school location (e.g., address, latitude/longitude), enrollment, and demographic characteristics. However, the vintage and location data in existing databases were found to be inadequate for this purpose. In some databases with high quality location data, some schools known to exist were not included (and schools known to be closed were included), while in others the location data quality was often limited and usually undocumented, etc. Therefore, various enhancements were made to a base set of school data, including additional data collection, review, development, documentation, and formatting, as described below.

```
  Time_Period_of_Content:
    Time_Period_Information:
      Single_Date/Time:
        Calendar_Date: July 2004
    Currentness_Reference: publication date
  Status:
    Progress: Complete
    Maintenance_and_Update_Frequency: None planned
  Spatial_Domain:
    Bounding_Coordinates:
      West_Bounding_Coordinate: -124.284769
      East_Bounding_Coordinate: -114.150000
      North_Bounding_Coordinate: 41.988100
      South_Bounding_Coordinate: 32.547787
  Keywords:
    Theme:
      Theme_Keyword: California
      Theme_Keyword: Schools
      Theme_Keyword: Public Schools
  Point_of_Contact:
    Contact_Information:
      Contact_Organization_Primary:
        Contact_Organization: EPA Office of Civil Rights
        Contact_Person: Loren Hall
      Contact_Address:
        Address_Type: mailing address
        Address: US Environmental Protection Agency (1201A)
        City: Washington
        State_or_Province: DC
        Postal_Code: 20460
        Country: USA
      Contact_Voice_Telephone: (202) 564-7289
      Contact_Facsimile_Telephone: (202) 501-1836
      Contact_Electronic_Mail_Address: hall.loren@epamail.epa.gov
      Hours_of_Service: 10:00 AM - 6:00 PM M-F
```

```
  Native_Data_Set_Environment: Microsoft Windows XP Version 5.1 (Build 2600)
Service Pack 1; ESRI ArcCatalog 9.0.0.535
Spatial_Data_Organization_Information:
  Direct_Spatial_Reference_Method: Vector
  Point_and_Vector_Object_Information:
    SDTS_Terms_Description:
      SDTS_Point_and_Vector_Object_Type: Entity point
      Point_and_Vector_Object_Count: 8472
Spatial_Reference_Information:
  Horizontal_Coordinate_System_Definition:
    Geographic:
      Latitude_Resolution: 0.000000
      Longitude_Resolution: 0.000000
      Geographic_Coordinate_Units: Decimal degrees
    Geodetic_Model:
      Horizontal_Datum_Name: North American Datum of 1983
      Ellipsoid_Name: Geodetic Reference System 80
      Semi-major_Axis: 6378137.000000
      Denominator_of_Flattening_Ratio: 298.257222
Entity_and_Attribute_Information:
  Detailed_Description:
    Entity_Type:
      Entity_Type_Label: Schools
    Attribute:
      Attribute_Label: FID
      Attribute_Definition: Internal feature number.
      Attribute_Definition_Source: ESRI
      Attribute_Domain_Values:
        Unrepresentable_Domain: Sequential unique whole numbers that are
automatically generated.
    Attribute:
      Attribute_Label: Shape
      Attribute_Definition: Feature geometry.
      Attribute_Definition_Source: ESRI
      Attribute_Domain_Values:
        Unrepresentable_Domain: Coordinates defining the features.
    Attribute:
      Attribute_Label: NCES_ID
      Attribute_Definition: National Center for Education Statistics (NCES)
nationally unique school identification code
    Attribute:
      Attribute_Label: SCHOOL_ID
      Attribute_Definition: State-wide 7-digit unique school identification
code
    Attribute:
      Attribute_Label: SCH_DIST
      Attribute_Definition: School district name
    Attribute:
      Attribute_Label: SCH_NAME
      Attribute_Definition: School name
    Attribute:
      Attribute_Label: SCH_ADD
      Attribute_Definition: School address
    Attribute:
      Attribute_Label: CITY_NAME
      Attribute_Definition: City name
    Attribute:
```

```
      Attribute_Label: ZIP_CODE
      Attribute_Definition: Zip code
    Attribute:
      Attribute_Label: ZIP_4
      Attribute_Definition: Zip Plus 4
    Attribute:
      Attribute_Label: PHONE
      Attribute_Definition: School phone number
    Attribute:
      Attribute_Label: GRD_ENROL
      Attribute_Definition: Grade span reported to the California Department
of Education for this school.
    Attribute:
      Attribute_Label: SCH_TYPE
      Attribute_Definition: The type of school.  As of publication date the
possible values were: alternative, Co Community (County Community), community
day, continuation, elementary, high school, junior high, juvenile hall, K-12,
middle, opportunity, or special Ed.  Additional values may include: other
county-wide programs, single elementary school in district, single high
school in district, state special school, or adult education center.
    Attribute:
      Attribute_Label: STAT0102
      Attribute_Definition: Status during 2001-2002 school year.  As of
publication date the possible values were:  open, closed, or retired.
Additional values may include: merged and pending.
    Attribute:
      Attribute_Label: LATITUDE
      Attribute_Definition: Latitude coordinate (decimal degrees)
    Attribute:
      Attribute_Label: LONGITUDE
      Attribute_Definition: Longitude coordinate (decimal degrees)
    Attribute:
      Attribute_Label: ZIP_LAT
      Attribute_Definition: Latitude of zip code centroid of school as
geocoded by TeleAtlas or GDT (decimal degrees)
    Attribute:
      Attribute_Label: ZIP_LON
      Attribute_Definition: Longitude of zip code centroid of school as
geocoded by TeleAtlas or GDT (decimal degrees)
    Attribute:
      Attribute_Label: DATUM
      Attribute_Definition: Datum code
    Attribute:
      Attribute_Label: REF_POINT
      Attribute_Definition: Reference point code
    Attribute:
      Attribute_Label: SCALE
      Attribute_Definition: Scale of source for coordinate capture
    Attribute:
      Attribute_Label: COL_MTH_CD
      Attribute_Definition: Code for how latitude/longitude was obtained
    Attribute:
      Attribute_Label: ACCURACY
      Attribute_Definition: Accuracy of latitude/longitude
    Attribute:
      Attribute_Label: SOURCE_DAT
```

```
      Attribute_Definition: Original source of latitude/longitude and other
school attribute data
    Attribute:
      Attribute_Label: SOURCE_LOC
      Attribute_Definition: Final source of school latitude/longitude
    Attribute:
      Attribute_Label: CONFIDENCE
      Attribute_Definition: Confidence of latitude/longitude
    Attribute:
      Attribute_Label: VERIF_METH
      Attribute_Definition: School location verification method
    Attribute:
      Attribute_Label: FEAT_TYPE
      Attribute_Definition: Geographic feature type.  As of publication date
the possible value was: point.  Additional values may include: line, polygon.
    Attribute:
      Attribute_Label: STATEFIPS
      Attribute_Definition: Federal Information Processing Standards (FIPS)
code for states, assigned by ICF Consulting
    Attribute:
      Attribute_Label: CNTYFIPS
      Attribute_Definition: County FIPS code, assigned by ICF Consulting
    Attribute:
      Attribute_Label: COUNTY
      Attribute_Definition: County name, assigned by ICF Consulting
    Attribute:
      Attribute_Label: CCD_NAME
      Attribute_Definition: Census county division name (populated ICF
Consulting using point-in-polygon operation on original school coordinates
from CCD and CDE)
    Attribute:
      Attribute_Label: TRACT90
      Attribute_Definition: 1990 Census tract from CCD
    Attribute:
      Attribute_Label: BLKGRP90
      Attribute_Definition: 1990 Census block group from CCD
    Attribute:
      Attribute_Label: PLACE
      Attribute_Definition: Census Place FIPS code for matched point from CCD
    Attribute:
      Attribute_Label: MCD
      Attribute_Definition: Minor Civil Division from CCD
    Attribute:
      Attribute_Label: MSA
      Attribute_Definition: Metropolitan Statistical Area from CCD
    Attribute:
      Attribute_Label: QUAD_ID
      Attribute_Definition: Corresponding U.S. Geological Survey (USGS)
1:24,000 Topographic Map Quadrangle ID used in verifying school location
based on Geographic Names Information System (GNIS)
    Attribute:
      Attribute_Label: COLL_DATE
      Attribute_Definition: Date latitude/longitude was obtained
    Attribute:
      Attribute_Label: COMMENT
      Attribute_Definition: General comments about coordinates and
verification
```

```
Distribution_Information:
  Resource_Description: Downloadable Data
  Standard_Order_Process:
    Digital_Form:
      Digital_Transfer_Information:
        Transfer_Size: 0.226
Metadata_Reference_Information:
  Metadata_Date: 20040715
  Metadata_Contact:
    Contact_Information:
      Contact_Organization_Primary:
        Contact_Organization: EPA Office of Civil Rights
        Contact_Person: Loren Hall
      Contact_Address:
        Address_Type: mailing address
        Address: US Environmental Protection Agency (1201A)
        City: Washington
        State_or_Province: DC
        Postal_Code: 20460
        Country: USA
      Contact_Voice_Telephone: (202) 564-7289
      Contact_Facsimile_Telephone: (202) 501-1836
      Contact_Electronic_Mail_Address: hall.loren@epamail.epa.gov
      Hours_of_Service: 10:00 AM - 6:00 PM M-F
  Metadata_Standard_Name: FGDC Content Standards for Digital Geospatial
Metadata
  Metadata_Standard_Version: FGDC-STD-001-1998
  Metadata_Time_Convention: local time
  Metadata_Extensions:
    Online_Linkage: http://www.esri.com/metadata/esriprof80.html
    Profile_Name: ESRI Metadata Profile
```

ATTACHMENT 3. PROJECT SPECIFIC DATA DOCUMENTATION (INVESTIGATION SENSITIVE)

This attachment describes a set of schools selected from the California public schools database for purposes of additional review and enhancement of the location information. This list has been developed in response to a specific complaint filed under Title VI of the Civil Rights Act of 1964 and EPA's implementing regulations.

The complaint investigation involves conducting an exposure and disparity analysis, which schools as the receptor locations for exposure. In these analyses, reasonably accurate school locations are important for achieving accurate results, so an effort was made to obtain acceptable location data for public schools in the state. However, not all schools' locations may be equally important in an assessment, depending on how close they are to usage sources, or other factors that may influence an assessment. A greater degree of effort was made for obtaining and verifying location data quality for these "priority" schools.

This attachment provides (1) background on the complaint, (2) the approach used to identify priority schools from the full database of California public schools and enhance these schools' location information, and (3) a description of the resulting data table fields.

Background

The subject complaint (number 16R-99-R9) was filed by the California Rural Legal Assistance Foundation *et al.* on behalf of children and parents who attend the following six California public schools:

1. Rio Plaza Elementary School, unincorporated Oxnard, Ventura County, California
2. Rio Mesa High School, unincorporated Oxnard, Ventura County, California
3. Pajaro Middle School, Pajaro, Monterey County, California
4. Barton Elementary School, Salinas, Monterey County, California
5. Macquiddy Elementary School, Watsonville, Santa Cruz County, California
6. Ohlone Elementary School, Pajaro, Monterey County, California

The complaint was brought concerning actions of the California Department of Pesticide Regulation (CDPR). CDPR is responsible for regulating the use of methyl bromide pursuant to the California Food and Agriculture Code and the California Code of Regulations, and is a recipient of federal financial assistance.

The Angelita C. complaint contains several allegations, some of which assert adverse disparate impacts on the complainants due to exposure to environmental stressors arising from the permitting of methyl bromide use in California, particularly in the vicinity of schools. All six schools have a high percentage of Hispanic ethnic children that are represented by the complainants. The complainants allege that California schoolchildren of color, primarily Latinos, are adversely impacted by agricultural use of methyl bromide in close proximity to California public schools, and that this adverse impact is disproportionately borne.

As described above, the complaint identified six specific schools of interest. The complaint also cited work conducted by the Environmental Working Group (EWG) that involved a series of studies of methyl bromide use in 1995 and 1998 near California schools. The most recent of these studies identifies over 400 schools that are within 1.5 miles of at least 100 pounds of methyl bromide applications in 1998 (see Methyl Bromide in California, http://www.ewg.org/california/ca-mb.html, last accessed November 7, 2003, and in particular the report, *An Ill Wind: Methyl Bromide Use Near California Schools*, February 2000).

Approximate Center of School Activity. Most of the school location coordinates in the full database relate to the front (street) area of the building, and were derived from address matching. For the complaint schools, however, the location coordinates have been determined based on the core group of school buildings and facilities in which most of the students are assumed to be throughout the day. Specifically, the core group of buildings were identified (e.g., from aerial photos). In some cases these areas were digitized, converted to a GIS polygon, and analyzed using ArcGIS to obtain the centroid of the polygon, while others were derived from inspection aerial photos.

Approach for Identifying Priority Schools and Enhancing Location Information

EPA conducted four steps to identify priority schools and enhance their location data:

(1) Start with the EWG schools;
(2) Identify schools within close proximity to the schools from the first step;
(3) Identify the largest schools;
(4) Obtaining adequate quality locations for all priority schools; and
(5) Verify location information for selected schools.

For the first step, EPA assigned priority status to virtually all of the EWG schools because of the potential proximity of these schools to agricultural areas of the state with known methyl bromide use (count=416; with all of the complaint schools also included as an EWG school). Two of the original 418 EWG records were excluded, one (Northwood High in Irvine) because it was a duplicate of another record and the other (Children's Center in Lodi) because it (1) was not found in the full schools database, (2) was noted by EWG as having an enrollment of zero in 1998, and (3) has an address that is virtually identical with another school in the database (Lawrence Elementary, at 721 vs.701 Calaveras Street, Lodi).

For the second step, EPA sought to identify additional schools that may be near relatively high methyl bromide usage areas of the state. These additional priority schools are defined as all California public schools that share the same Census county division (a Census-defined subdivision of counties[45]) as an EWG school (count=3,961).

The third step involved ensuring that all of the largest schools in the state (as measured by total enrollment) were identified.. These schools were defined at the top ten percent in terms of enrollment (regardless of ethnicity; count=860). These schools, many of which are not in or near

[45] For a complete definition of Census county divisions, see Cartographic Boundary Files: County Subdivisions at http://www.census.gov/geo/www/cob/cs_metadata.html (last accessed on November 7, 2003). For an example map and additional detail on Census terminology, see Small-Area Geography Overview at http://www.census.gov/mso/www/rsf/geo_con/index.htm (last accessed on November 7, 2003).

agricultural areas, require accurate information because of their large student populations and thus potential influence on the investigation's disparity analyses.

The fourth step involved reviewing the list of priority schools to ensure that each had a location of at lease satisfactory quality, such as those resulting from block number address matching. Ultimately, this process was also performed for almost all schools in the state which were near agricultural areas or had more than a few students. The procedures involved using alternative address match sources, such as the online Maporama site, looking up location data for schools from the USGS Geographic Names Information System (GNIS) database, and other methods such as examining aerial photos of the nearby general area for sites with recognizable features such as groups of buildings combined with recreational facilities such as tracks and fields. For the complaint schools, a special procedure was used to outline (digitize) the area of the school grounds containing buildings, and thus most likely to be where student activities occur. From this area, a point centroid was calculated to use as the location for these schools.

The fifth step, verifying location information, coincided with the effort described in Section 3.3 of the main text to target location verification of selected schools. Thus, a subset of the priority schools (count=196) were verified using a variety of methods, including comparison with the GNIS location, aerial photos, and GPS. This subset includes (1) over 100 schools that constitute the 25 percent of schools listed by EWG as being near the highest applications of methyl bromide in 1998, (2) approximately 70 random priority schools, (3) 7 schools at which an approximation was made of the center of student activity (primarily the complaint schools; see box at right), and (4) schools identified as having potentially high levels of exposure in pilot modeling analyses.

Priority Schools Data Table

Table ATT3-1 describes the data table added to investigation sensitive version of the California public schools database.

Table Att3-1. Priority Schools Table
Table Name: Priority
Source: ICF Consulting, Inc.
Description: This relate table contains one record with flag fields (Y or blank) for any school from the school database that was considered to be a "priority" school for location coordinate verification. Priority schools include (1) EWG schools (count=408), (2) schools that share census county division with an EWG school (count=3,961), and (3) one of the top approximately ten percent highest school population schools (count=860).

Field Name	Description
NCES_ID	NCES nationally unique school identification code
SCHOOL_ID	State-wide 7-digit unique School identification code
HIPOP	High population school (top 10% highest populations in school data set, regardless of race/ethnicity)
EWG_SCHOOL	School in EWG list
EWG_CCD	School in same Census county division as a school in the EWG list

Complaint School Location Information

This section addresses the six complaint schools listed above. For each school, Table Att3-2 provides the mailing address, a list of various website links related to the school, information provided by the California Air Resources Board (CARB) about the school, each school's U.S. Department of Education Common Core of Data (CCD) coordinates supplemented by the Agency for Toxic Substances and Disease Registry (ATSDR), and final estimated activity center coordinates as described above.

The first photo presented below for each of these schools shows the following: (1) the location of the CCD-supplemented coordinates, assuming datum NAD27 (⚠; see Section 3.3 for a description of this issue); (2) the final estimated activity center (⊙; datum NAD83); and (3) the core group of school buildings and facilities used for the activity center derivation

Table Att3-2. Location Information for Complaint Schools

School name	Address	URLs	Site as described by CARB	CCD-coordinates (decimal deg)[a]	Final coordinates (decimal-deg)[b]
Barton (Virginia Rocca) Elementary School	680 Las Casitas Dr., Salinas, CA 93905	http://www.greatschools.org/california/watsonville/5748-Pajaro-Middle-School/ http://www.alisal.org/apps/pages/index.jsp?uREC_ID=41930&type=d	Not available	Lat : 36.6949 Lon -121.61327	Lat : 36.6957032093 Lon : -121.61566379
MacQuiddy Elementary School	330 Martinelli St., Watsonville, CA 95076-2810	http://www.greatschools.org/search/search.page?search_type=0&q=MacQuiddy+Elementary&state=CA&c=school http://www.pvusd.net/schools/south_zone/macquiddy_elementary_school/index.html	MacQuiddy Elementary School, 331 Martinelly St., Watsonville, CA 95076-2810, John McCann Env. Helath & Safety Officer (831) 728-6248 ext 291 Swction/Township/Range: S.33/T.11E/R.2E Lat: 36.92662445710 Lon: -121.75176910500	Lat: 36.92641 Lon: -121.74986	Lat: 36.8703470884 Lon: -121.75171657
Ohlone Elementary School	21 Bay Farms Rd, Watsonville, CA 95076	http://www.greatschools.org/search/search.page?search_type=0&q=MacQuiddy+Elementary&state=CA&c=school http://www.pvusd.net/schools/south_zone/ohlone_elementary_school/index.html	Not available	Lat: 36.8718 Lon: -121.7471	Lat: 36.8703470884 Lon: -121.76182503
Pajaro Middle School	250 Salinas Road, Watsonville, CA 95076	http://www.greatschools.org/california/watsonville/5748-Pajaro-Middle-School/ http://www.pvusd.net/schools/south_zone/pajaro_middle_school/index.html	Pajaro Middle School, 250 Salinas Road, Watsonville, CA 95076, Jackie Defendis Principal (831) 728-6238, Section/Township/Range: S.9/T.12S/R.2E Lat: 36.89910468310 Lon: -121.64856392000	Lat: 36.90078 Lon: -121.7471	Lat: 36.899171158 Lon: -121.74881833
Rio Mesa High School	545 Central Ave,. Oxnard CA 93030	http://www.greatschools.org/california/oxnard/7657-Rio-Mesa-High-School/ http://www.riomesahigh.us/	Not available	Lat: 34.25882 Lon: -119.15322	Lat: 34.2535787651 Lon: -119.14443181
Rio Plaza Elementary School	600 Simon Way, Oxnard CA 93030	http://www.greatschools.org/california/oxnard/7693-Rio-Plaza-Elementary-School/ http://www.rio.k12.ca.us/Rio%20Plaza/	Not available	Lat: 34.24009 Lon: -119.157	Lat: 34.2397197895 Lon: -119.15844699

[a] Likely based on datum NAD27
[b] Based on datum NAD83

Figure Att3-1A. Aerial Photo with Mapped Coordinates: Barton (Virginia Rocca) Elementary School

Figure Att3-1B. Additional Aerial Photo: Barton (Virginaia Rocca) Elementary School

Figure Att3-2A. Aerial Photo with Mapped Coordinates: MacQuiddy Elementary School

Figure Att3-2B. Additional Aerial Photo: MacQuiddy Elementary School

Figure Att3-3A. Aerial Photo with Mapped Coordinates: Ohlone Elementary School

Figure Att3-3B. Additional Aerial Photo: Ohlone Elementary School

Figure Att3-4A. Aerial Photo with Mapped Coordinates: Pajaro Middle School

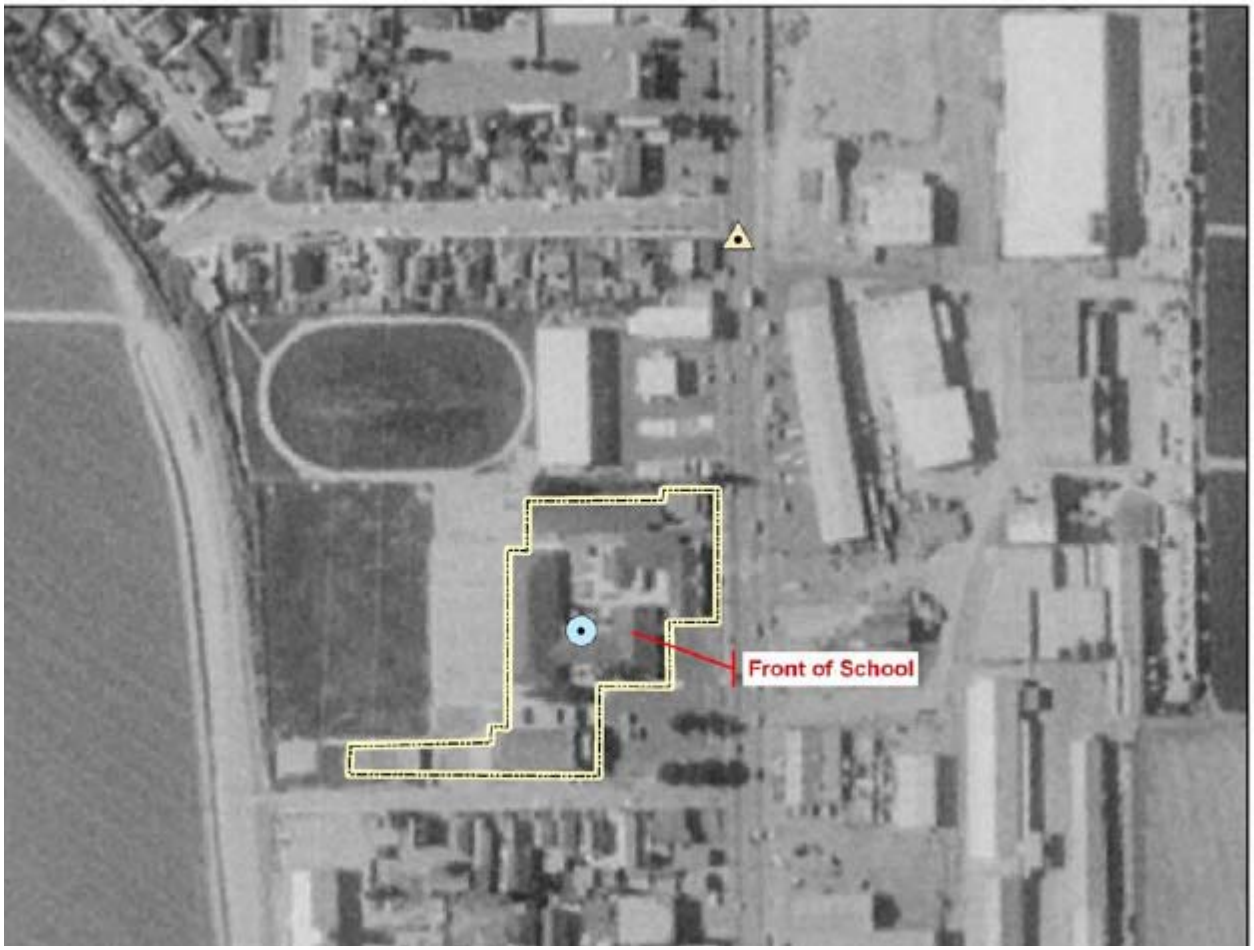


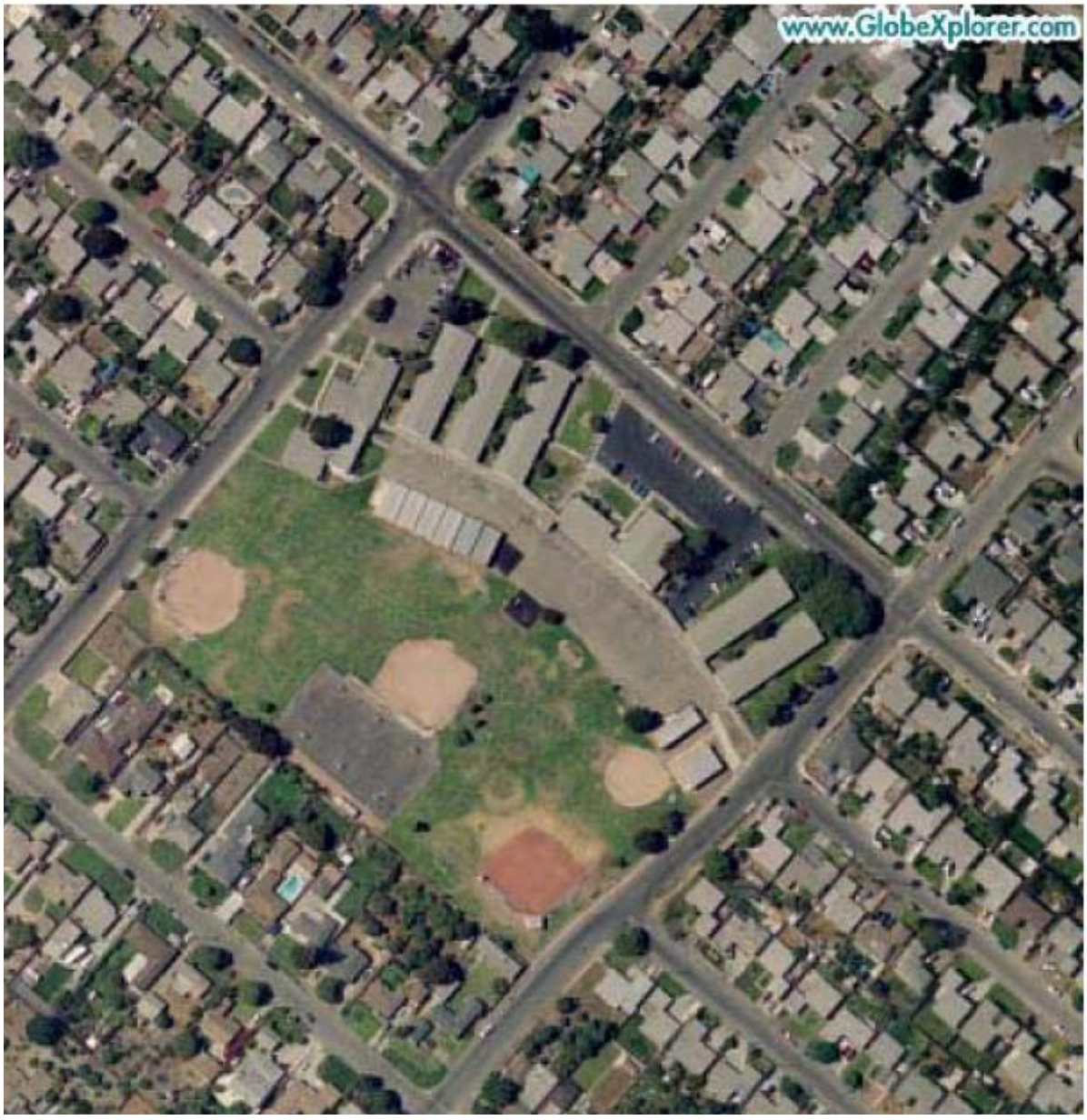
Figure Att3-4B. Additional Aerial Photo: Pajaro Middle School

Figure Att3-5A. Aerial Photo with Mapped Coordinates: Rio Mesa High School

Figure Att3-5B. Additional Aerial Photo: Rio Mesa High School

Figure Att3-6A. Aerial Photo with Mapped Coordinates: Rio Plaza Elmentary School

Figure Att3-6B. Additional Aerial Photo: Rio Plaza Elementary School



APPENDIX I. FARMLAND COMPOSITE AML PROGRAM SERIES

I. The main program that calls all other programs (Master.aml)

```
/*************************************************************************
/* [investigation sensitive portion of title deleted] - California Schools
Project
/* Created by: Matt Werth, ICF Consulting GIS Dept
/* 11/20/2003
/*
/* This is the master AML to recreate the farmland dataset from the original
/* sources.
/*
/*************************************************************************

/* process DOC
/* convert the source SHP data into coverage
/* aggregate classes into Farm, NotFarm, and delete Not Mapped
/* dissolve, reproject, and build
&r process_doc.aml

&r DelSub2Acre.aml farmdoc

/* process DOW
/* convert the source SHP data into coverage
/* aggregate classes into Farm, NotFarm.  Not Mapped was already removed from
source.
/* assign splatter along county boundary as NotFarm
/* dissolve, reproject, and build
&r process_dow.aml

&r DelSub2Acre.aml farmdow

/* process FRAP
/* convert the source SHP data into coverage
/* build
&r process_frap.aml

/* process Santa Cruz
/* convert the source SHP data into coverage
/* build
&r process_sc.aml

/* perform stacking via UNION
/* establish CLASS as 'FarmDoc',
'NotFarmDoc','FarmDow','NotFarmDow','FarmFrap'
/* dissolve by CLASS
&r mergecov.aml

/* artifacts are: DOW and FRAP polygons that are
/*                - smaller than 5 acres (size is easy to change)
/*                - adjacent to (DOC, NotDOC, DOW, NotDOW)
/*                - adjacent to nothing (distinguish outer edge from interior
donut holes)
```

```
/* identify artifacts, save into intermediate coverage, delete from current
coverage
&r rm_artifacts.aml

/* use SC cookie cutter to erase a hole in existing farmland composite
/* union in SantaCruz fields (union so we retain PMTSITEID field)
&r MergeSantaCruz.aml

/* Final Cleanup and packaging
&r FinalCleanup.aml
```

II. Program where FMMP data are preprocessed prior to union with other sources (Process_doc.aml)

```
&if [exists farmdoc_rgn -cover] &then
  kill farmdoc_rgn all
&if [exists cls -info] &then
  killinfo cls

shapearc SourceInput/farmdoc.shp farmdoc_rgn cls

projectdefine cover farmdoc_rgn
Projection               ALBERS
Units                    METERS
Datum                     NAD27
Spheroid             CLARKE1866
Parameters
34  0  0.000
40 30  0.000
-120  0  0.00
0  0  0.000
0.00000
-4000000.0000


&if [exists farmdoc_cn -cover] &then
kill farmdoc_cn all

clean farmdoc_rgn farmdoc_cn # 0.2
createlabels farmdoc_cn
build farmdoc_cn

&if [exists farmdoc_pol -cover] &then
kill farmdoc_pol all
&if [exists clss -info] &then
killinfo clss

regionpoly farmdoc_cn farmdoc_pol cls clss


arcedit
edit farmdoc_pol poly
additem class_doc 15 15 c
/* there are three sliver holes surrounded by NotFarmDoc in the source
dataset
```

```
/* therefore we can just assign them as NotFarmDoc so that they get dissolved
away
sel polygon_ty = ''
&s num = [show number select]
&if %num% <> 0 &then
  calc class_doc = 'NotFarmDoc'

sel polygon_ty = 'P' or polygon_ty = 'S' or polygon_ty = 'U' or polygon_ty =
'L'
asel polygon_ty = 'I' or polygon_ty = 'N' or polygon_ty = 'LP'
calc class_doc = 'FarmDoc'
sel polygon_ty = 'Z'
calc class_doc = 'NotMappedDoc'
sel class_doc = ''
calc class_doc = 'NotFarmDoc'

/*************************
/* Delete Not Mapped
/* note we must do this after NotFarmDoc is asssigned
sel class_doc = 'NotMappedDoc'
delete

save
quit

&if [exists farmdoc_ds -cover] &then
kill farmdoc_ds all

dissolve farmdoc_pol farmdoc_ds class_doc

&if [exists farmdoc_alb -cover] &then
kill farmdoc_alb all

project cover farmdoc_ds farmdoc_alb
output
Projection               ALBERS
Datum                     NAD83
Units                      FEET
Spheroid                GRS1980
Parameters
 32  0  0.000
 42  0  0.000
 -120  0  0.00
 30  0  0.000
 0.00000
 0.00000
end

build farmdoc_alb

/*****************************
/* clean up
&if [exists farmdoc_rgn -cover] &then
  kill farmdoc_rgn all

&if [exists farmdoc_cn -cover] &then
  kill farmdoc_cn all
```

```
&if [exists clss -info] &then
  killinfo clss

&if [exists farmdoc_pol -cover] &then
  kill farmdoc_pol all

&if [exists farmdoc_ds -cover] &then
  kill farmdoc_ds all
```

III. Program where LUSD data are preprocessed prior to union with other sources (Process_dow.aml)

```
&if [exists farmdow_rgn -cover] &then
kill farmdow_rgn all

shapearc SourceInput/farmdow farmdow_rgn cls

projectdefine cover farmdow_rgn
Projection           TRANSVERSE
Datum                     NAD27
Units                    METERS
Spheroid             CLARKE1866
Parameters
0.99960000
-120  0  0.00
0  0  0.000
500000.00000
0.00000

&if [exists farmdow_cn -cover] &then
kill farmdow_cn all

clean farmdow_rgn farmdow_cn # 0.4
createlabels farmdow_cn
build farmdow_cn

&if [exists clss -info] &then
killinfo clss

&if [exists farmdow_pol -cover] &then
kill farmdow_pol all

regionpoly farmdow_cn farmdow_pol cls clss

Arcedit
edit farmdow_pol poly
additem class_dow 15 15 c

/* Select the appropriate polygons and classify them into FarmDow, Not
Farmdow, and Not Mapped

/*************************
/* Farmland
```

```
select class1 = 'G' or  class1 = 'R' or  class1 = 'F' or  class1 = 'T' or
class1 = 'D' or  class1 = 'C' or  class1 = 'V' or  class1 = 'I'
asel class1 = 'P' and subclass1 = '7'
calc class_dow = 'FarmDow'
/*************************
/* Not Farmland
sel class1 = 'P' and subclass1 <> '7'
aselect class1 = 'S' or  class1 = 'U' or  class1 = 'N' or class1 = 'NW' or
class1 = 'NB' or class1 = 'NR' or class1 = 'NV'
aselect class1 = 'UC' or  class1 = 'UI' or  class1 = 'UL' or class1 = 'UR' or
class1 = 'UV'
calc class_dow = 'NotFarmDow'
sel class1 = 'NS' or  class1 = 'E' or  class1 = 'Z'
&if [show number select] > 0 &then
  calc class_dow = 'NotMappedDow'

/* correct the null garbage around the county boundary
select class_dow = ''
calc class_dow = 'NotFarmDow'

/*************************
/* Delete Not Mapped
sel class_dow = 'NotMappedDow'
&if [show number select] > 0 &then
  calc class_dow = 'NotMappedDow'
save
quit

/*Then used dissolve command to get rid of the lines boardering polygons of
the same class.

&if [exists farmdow_ds -cover] &then
  kill farmdow_ds all

dissolve farmdow_pol farmdow_ds class_dow

&if [exists farmdow_alb -cover] &then
  kill farmdow_alb all

project cover farmdow_ds farmdow_alb
output
Projection               ALBERS
Datum                     NAD83
Units                      FEET
Spheroid                GRS1980
Parameters
 32  0  0.000
 42  0  0.000
 -120  0  0.00
 30  0  0.000
 0.00000
 0.00000
end

build farmdow_alb
```

```
/*****************************
/* clean up
&if [exists farmdow_rgn -cover] &then
  kill farmdow_rgn all

&if [exists farmdow_cn -cover] &then
  kill farmdow_cn all

&if [exists clss -info] &then
  killinfo clss

&if [exists farmdow_pol -cover] &then
  kill farmdow_pol all

&if [exists farmdow_ds -cover] &then
  kill farmdow_ds all
```

IV. Program where FRAP data are preprocessed prior to union with other sources (Process_frap.aml)

```
&if [exists tmpFrap1 -grid] &then
   kill tmpFrap1 all

/* FRAP AGRICULTURAL VALUES (4,6,7,8,9,10,11,12)
GRID
  tmpFrap1 = con(SourceInput/FRAP/fveg02_2g >= 4 AND ~
                 SourceInput/FRAP/fveg02_2g <= 12 AND ~
                 SourceInput/FRAP/fveg02_2g <> 5, ~
                 1)
quit

&if [exists tmpFrap2 -cover] &then
  kill tmpFrap2 all

/* convert grid to polygon cov
gridpoly tmpFrap1 tmpFrap2

projectdefine cover tmpFrap2
Projection           ALBERS
Datum                NAD27
Units                METER
Spheroid             CLARKE1866
Parameters
34 0 0
40 30 0
-120 0 0
0 0 0
0.00000
-4000000.00000
end

&if [exists tmpFrap3 -cover] &then
  kill tmpFrap3 all

project cover tmpFrap2 tmpFrap3
```

```
output
Projection               ALBERS
Datum                     NAD83
Units                      FEET
Spheroid                GRS1980
Parameters
 32  0  0.000
 42  0  0.000
 -120  0  0.00
 30  0  0.000
 0.00000
 0.00000
end

build tmpFrap3

additem tmpFrap3.pat tmpFrap3.pat class_frap 15 15 C

tables
select tmpFrap3.pat
reselect grid-code = 1
calc class_frap = 'FarmFrap'
quit

&if [exists tmpFrap4 -cover] &then
  kill tmpFrap4 all

/* Our artifact routine deletes the artifacts below 10 acres but above ten
acres there
/* are a mixed bag of keepers and legit artifacts.  We manually identified
and saved the
/* artifacts to be deleted into the clip coveargе that we can remove ahead of
time.
erase tmpFrap3 artifact_man tmpFrap4

&if [exists farmfrap_alb -cover] &then
  kill farmfrap_alb all


/* dissolve it
dissolve tmpFrap4 farmfrap_alb class_frap

build farmfrap_alb

/*****************************
/* clean up
&if [exists tmpFrap1 -grid] &then
   kill tmpFrap1 all

&if [exists tmpFrap2 -cover] &then
  kill tmpFrap2 all

&if [exists tmpFrap3 -cover] &then
  kill tmpFrap3 all

&if [exists tmpFrap4 -cover] &then
  kill tmpFrap4 all
```

V. Program where Santa Cruz data are preprocessed prior to union with other sources (Process_sc.aml)

```
&&if [exists farmsc_rgn -cover] &then
  kill farmsc_rgn all

shapearc SourceInput/sc_fields farmsc_rgn cls


projectdefine cover farmsc_rgn
Projection           Albers
Datum                NAD83
Units                feet
Spheroid             GRS1980
Parameters
32 0 0
42 0 0
-120 0 0
30 0 0
0.00000
0.00000


&if [exists farmsc_cn -cover] &then
  kill farmsc_cn all

clean farmsc_rgn farmsc_cn # 0.4
createlabels farmsc_cn
build farmsc_cn

&if [exists clss -info] &then
killinfo clss

&if [exists farmsc_alb -cover] &then
kill farmsc_alb all

regionpoly farmsc_cn farmsc_alb cls clss

/* dissolving is not needed.

build farmsc_alb



/*****************************
/* clean up
&if [exists farmsc_rgn -cover] &then
  kill farmsc_rgn all

&if [exists farmsc_cn -cover] &then
  kill farmsc_cn all

&if [exists clss -info] &then
  killinfo clss
```

VI. Program that handles the removal of sub-two acre farmland polygons (DelSub2Acre.aml)

```
/* w h:\ca_schools\rebuild_farmland_jue\work
/* To run this progam, type &run del_sub2acre.aml farmdoc
/* or &run DelSub2Acre.aml farmdow

&args cov

&if [exists %cov% -cover] &then
kill %cov% all
copy %cov%_alb %cov%

&if %cov% = farmdow &then
  &do
  &s itm1 class_dow
  &s val1 FarmDow
  &s val1 = [quote %val1%]
  &s val2 NotFarmDow
  &s val2 = [quote %val2%]
  &end
&else &if %cov% = farmdoc &then
  &do
  &s itm1  class_doc
  &s val1  FarmDoc
  &s val1 = [quote %val1%]
  &s val2  NotFarmDoc
  &s val2 = [quote %val2%]
  &end
&else
  &return "Invalid parameter sent.  Expecting 'farmdoc' or 'farmdow'"




&s artif artif_num
&s donut dnt_num


relate add
left
%cov%.pat
info
lpoly#
%cov%#
ordered
rw
right
%cov%.pat
info
rpoly#
%cov%#
```

```
ordered
rw
~

build %cov% arc

&if [show program] <> 'ARCEDIT' &then
arcedit
edit %cov% arc

/* Selecting the arcs boardering FarmFrap and universal polygon and write the
polygon number
/* of FarmFrap into the text file.

&s file1 = [open %artif% opens -w]

&type Selecting the arcs

sel left//%itm1% = %val1% and left//area < 87120 and right//%itm1% = %val2%
asel left//%itm1% = %val2% and left//area < 87120 and right//%itm1% = %val1%
&s num_sel1 = [show number select]
&if %num_sel1% > 0 &then
&do i = 1 &to %num_sel1%
&s num = [show select %i%]
&ty %num%
&s lpoly_num = [show arc %num% item lpoly#]
&ty left polygon number is %lpoly_num%
&s writestat = [write %file1% %lpoly_num%]
&end

&type Selecting the arcs
sel right//%itm1% = %val1% and right//area < 87120 and left//%itm1% = %val2%
asel right//%itm1% = %val2% and right//area < 87120 and left//%itm1% = %val1%


&s num_sel2 = [show number select]
&if %num_sel2% > 0 &then
&do i = 1 &to %num_sel2%
&s num = [show select %i%]
&ty num is %num%
&s rpoly_num = [show arc %num% item rpoly#]
&ty rpoly_num is %rpoly_num%
&s writestat = [write %file1% %rpoly_num%]
&end

&s total_num = %num_sel1% + %num_sel2%
&s closestat [close %file1% ]
&if %closestat% = 0 &then
&ty %file1% is successfully closed.
quit

&if [iteminfo %cov%.pat -info bd_in -exists] &then
dropitem %cov%.pat %cov%.pat bd_in

&type adding an item to hold the character of the frap polygon
additem %cov%.pat %cov%.pat bd_in 4 4 c
```

```
arcedit
edit %cov% poly

/* selecting all of the small polygons either boardering or floating

&s file1 = [open %artif% openstat -r]
&do i = 1 &to %total_num%
&s poly_num = [read %file1% readstat]
&if %i% = 1 &then
sel %cov%# = %poly_num%
&else
asel %cov%# = %poly_num%
&end
calc bd_in = 'IN'
&s closestat [close %file1% ]
save
quit



relate add
left
%cov%.pat
info
lpoly#
%cov%#
ordered
rw
right
%cov%.pat
info
rpoly#
%cov%#
ordered
rw
~


arcedit
edit %cov% arc
&s file2 = [open %donut% opens -w]

&type Selecting the arcs

sel left//%itm1% = %val1% and left//area < 87120 and right//%itm1% = ''
asel left//%itm1% = %val2% and left//area < 87120 and right//%itm1% = ''
&s num_sel1 = [show number select]
&if %num_sel1% > 0 &then
&do i = 1 &to %num_sel1%
&s num = [show select %i%]
&ty %num%
&s lpoly_num = [show arc %num% item lpoly#]
&ty left polygon number is %lpoly_num%
&s writestat = [write %file2% %lpoly_num%]
&end
```

```
&type Selecting the arcs
sel right//%itm1% = %val1% and right//area < 87120 and left//%itm1% = ''
asel right//%itm1% = %val2% and right//area < 87120 and left//%itm1% = ''
&s num_sel2 = [show number select]
&if %num_sel2% > 0 &then
&do i = 1 &to %num_sel2%
&s num = [show select %i%]
&ty num is %num%
&s rpoly_num = [show arc %num% item rpoly#]
&ty rpoly_num is %rpoly_num%
&s writestat = [write %file2% %rpoly_num%]
&end

&s total_num2 = %num_sel1% + %num_sel2%
&s closestat [close %file2% ]
&if %closestat% = 0 &then
&ty %file2% is successfully closed.

ef poly

/* Reselecting only the boardering polygons

&s file2 = [open %donut% openstat -r]
&do i = 1 &to %total_num2%
&s poly_num = [read %file2% readstat]
&if %i% = 1 &then
sel %cov%# = %poly_num%
&else
asel %cov%# = %poly_num%
&end
resel bd_in = 'IN'
&if [show number select] > 0 &then
calc bd_in = 'BD'

select area > 0 and area < 87120 and bd_in = ''
calc bd_in = 'FL'
save
&s closestat [close %file2% ]

select bd_in = 'FL' or bd_in = 'IN' or bd_in = 'BD'

&if [exists %cov%_sub2 -cover] &then
&do
arcplot
arc kill %cov%_sub2  all
quit
&end

/* saved these polygons into an intermediate file so we can QA it
put %cov%_sub2

select bd_in = 'FL'
delete

select bd_in = 'IN' and %itm1% = %val1%
aselect bd_in = 'BD' and %itm1% = %val1%
```

```
calc %itm1% = 'TEMP'
select bd_in = 'IN' and %itm1% = %val2%
aselect bd_in = 'BD' and %itm1% = %val2%
calc %itm1% = %val1%

select bd_in = 'IN' and %itm1% = 'TEMP'
aselect bd_in = 'BD' and %itm1% = 'TEMP'

calc %itm1% = %val2%

save

quit

build %cov%_sub2

&if [exists %cov%_2acre -cover] &then
  kill %cov%_2acre all

dissolve %cov% %cov%_2acre %itm1%

&if [exists %cov% -cover] &then
  kill %cov% all
```

VII. Program that handles merging of all farmland data sources (Mergecov.aml)

```
/* copy ../farmdoc/farmdoc_alb farmdoc
/* copy ../farmdow/farmdow_alb farmdow
/* copy ../farmfrap/farmfrap_pol farmfrap
/*copy ../santacruz/farmsc_pol farmsc

&if [exists fm_c_w -cover] &then
    kill fm_c_w all

union farmdoc_2acre farmdow_2acre fm_c_w

&if [exists fm_c_w_f -cover] &then
    kill fm_c_w_f all

union fm_c_w farmfrap_alb fm_c_w_f

arcedit
edit fm_c_w_f poly
additem class 15 15 c
sel class_doc = 'FarmDoc' or class_doc = 'NotFarmDoc'
calc class = class_doc
sel class = ''
resel class_dow = 'FarmDow' or class_dow = 'NotFarmDow'
calc class = class_dow
sel class = ''
resel class_frap = 'FarmFrap'
calc class = 'FarmFrap'
save
quit
```

```
/*****************************
/* clean up
&if [exists fm_c_w -cover] &then
    kill fm_c_w all
```

VIII. Program that handles merge of Santa Cruz polygons and attributes (MergeSantaCruz.aml)

```
&if [exists fm_c_w_f_nsc -cover] &then
  kill fm_c_w_f_nsc all


/* erase santa cruz cookie cutter
erase farm_nafacts  SourceInput/sc_cutter fm_c_w_f_nsc


&if [exists fm_c_w_f_sc -cover] &then
  kill fm_c_w_f_sc all

/* Union in Sanata Cruz
/* do union so that retain the PMTSITEID col from santa cruz
union fm_c_w_f_nsc farmsc_alb fm_c_w_f_sc

tables
select fm_c_w_f_sc.pat
reselect FARMSC_ALB# > 1
calc class = 'Santa Cruz'
quit

/*****************************
/* clean up
&if [exists fm_c_w_f_nsc -cover] &then
    kill fm_c_w_f_nsc all
```

IX. Program that handles removing merging artifacts (rm_artifacts.aml)

```
/* w c:\projects\ca_schools\rebuild_farmland_jue\work
/* w h:\ca_schools\rebuild_farmland_jue\work

/*&s %lngArtifactAcres% 87120  /* 2 acres
/*&s %lngArtifactAcres% 217800 /* 5 acres
&s lngArtifactAcres 435600   /* 10 acres
/*&s lngArtifactAcres 1089000   /* 25 acres
/*&s lngArtifactAcres 2178000   /* 50 acres
/*&s lngArtifactAcres 8712000   /* 200 acres

&if [exists  fm_c_w_f_dis -cover] &then
  kill fm_c_w_f_dis all

dissolve fm_c_w_f fm_c_w_f_dis class

build fm_c_w_f_dis arc
```

```
&s polynumfil frap_num
&s donut donut_num


relate add
left
FM_C_W_F_DIS.pat
info
lpoly#
FM_C_W_F_DIS#
ordered
rw
right
FM_C_W_F_DIS.pat
info
rpoly#
FM_C_W_F_DIS#
ordered
rw
~

&if [show program] <> 'ARCEDIT' &then
arcedit
edit FM_C_W_F_DIS arc

/* Selecting the arcs boardering FarmFrap and universal polygon and write the
polygon number
/* of FarmFrap into the text file.

&s file1 = [open %polynumfil% opens -w]

&type Selecting the arcs

sel left//class = 'FarmFrap' and left//area < %lngArtifactAcres%  and
right//class = 'FarmDoc'
asel left//class = 'FarmFrap' and left//area < %lngArtifactAcres%  and
right//class = 'FarmDow'
asel left//class = 'FarmFrap' and left//area < %lngArtifactAcres%  and
right//class = 'NotFarmDoc'
asel left//class = 'FarmFrap' and left//area < %lngArtifactAcres%  and
right//class = 'NotFarmDow'
asel left//class = 'FarmDow' and left//area < %lngArtifactAcres%  and
right//class = 'FarmDoc'
asel left//class = 'FarmDow' and left//area < %lngArtifactAcres%  and
right//class = 'NotFarmDoc'
&s num_sel1 = [show number select]
&if %num_sel1% > 0 &then
&do i = 1 &to %num_sel1%
&s num = [show select %i%]
&ty %num%
&s lpoly_num = [show arc %num% item lpoly#]
&ty left polygon number is %lpoly_num%
&s writestat = [write %file1% %lpoly_num%]
&end

&type Selecting the arcs
```

```
sel right//class = 'FarmFrap' and right//area < %lngArtifactAcres%  and
left//class = 'FarmDoc'
asel right//class = 'FarmFrap' and right//area < %lngArtifactAcres%  and
left//class = 'FarmDow'
asel right//class = 'FarmFrap' and right//area < %lngArtifactAcres%  and
left//class = 'NotFarmDoc'
asel right//class = 'FarmFrap' and right//area < %lngArtifactAcres%  and
left//class = 'NotFarmDow'
asel right//class = 'FarmDow' and right//area < %lngArtifactAcres%  and
left//class = 'FarmDoc'
asel right//class = 'FarmDow' and right//area < %lngArtifactAcres%  and
left//class = 'NotFarmDoc'


&s num_sel2 = [show number select]
&if %num_sel2% > 0 &then
&do i = 1 &to %num_sel2%
&s num = [show select %i%]
&ty num is %num%
&s rpoly_num = [show arc %num% item rpoly#]
&ty rpoly_num is %rpoly_num%
&s writestat = [write %file1% %rpoly_num%]
&end

&s total_num = %num_sel1% + %num_sel2%
&s closestat [close %file1% ]
&if %closestat% = 0 &then
&ty %file1% is successfully closed.
quit


&type adding an item to hold the character of the frap polygon
additem FM_C_W_F_DIS.pat FM_C_W_F_DIS.pat bd_in 4 4 c


arcedit
edit FM_C_W_F_DIS poly

/* selecting all of the Frap polygons that either boardering or inside
'FarmDoc'

&s file1 = [open %polynumfil% openstat -r]
&do i = 1 &to %total_num%
&s poly_num = [read %file1% readstat]
&if %i% = 1 &then
sel FM_C_W_F_DIS# = %poly_num%
&else
asel FM_C_W_F_DIS# = %poly_num%
&end
calc bd_in = 'IN'
&s closestat [close %file1% ]
save
quit
```

```
relate add
left
FM_C_W_F_DIS.pat
info
lpoly#
FM_C_W_F_DIS#
ordered
rw
right
FM_C_W_F_DIS.pat
info
rpoly#
FM_C_W_F_DIS#
ordered
rw
~


arcedit
edit FM_C_W_F_DIS arc
&s file2 = [open %donut% opens -w]

&type Selecting the arcs

sel left//class = 'FarmFrap' and left//area < %lngArtifactAcres%  and
right//class = ''
asel left//class = 'FarmDow' and left//area < %lngArtifactAcres%  and
right//class = ''
&s num_sel1 = [show number select]
&if %num_sel1% > 0 &then
&do i = 1 &to %num_sel1%
&s num = [show select %i%]
&ty %num%
&s lpoly_num = [show arc %num% item lpoly#]
&ty left polygon number is %lpoly_num%
&s writestat = [write %file2% %lpoly_num%]
&end

&type Selecting the arcs
sel right//class = 'FarmFrap' and right//area < %lngArtifactAcres%  and
left//class = ''
asel right//class = 'FarmDow' and right//area < %lngArtifactAcres%  and
left//class = ''
&s num_sel2 = [show number select]
&if %num_sel2% > 0 &then
&do i = 1 &to %num_sel2%
&s num = [show select %i%]
&ty num is %num%
&s rpoly_num = [show arc %num% item rpoly#]
&ty rpoly_num is %rpoly_num%
&s writestat = [write %file2% %rpoly_num%]
&end

&s total_num2 = %num_sel1% + %num_sel2%
&s closestat [close %file2% ]
&if %closestat% = 0 &then
&ty %file2% is successfully closed.
```

```
ef poly

/* Reselecting only the boardering polygons

&s file2 = [open %donut% openstat -r]
&do i = 1 &to %total_num2%
&s poly_num = [read %file2% readstat]
&if %i% = 1 &then
sel FM_C_W_F_DIS# = %poly_num%
&else
asel FM_C_W_F_DIS# = %poly_num%
&end
&s closestat [close %file2% ]

resel bd_in = 'IN'
calc bd_in = 'BD'
&if [exists artifacts -cover] &then
  &do
  arcplot
  arc kill artifacts all
  quit
  &end

&if [show number select] > 0 &then
&do
put artifacts
delete
&end

save
quit

build artifacts


/* disolve the interior null donut holes that were created when the artifacts
we deleted
/* were interior and thus replace with a null polygon.
&if [exists  farm_nafacts -cover] &then
  kill farm_nafacts all

dissolve fm_c_w_f_dis farm_nafacts class

&if [exists fm_c_w_f_dis -cover] &then
  kill fm_c_w_f_dis all

&type Program run successfully!
```

X. Program that removes unnecessary fields and adds field for FarmID (FinalCleanup.aml)

```
/*************************************************************************
```

```
/* [investigation sensitive portion of title deleted] - California Schools
Project
/* Created by: Matt Werth, ICF Consulting GIS Dept
/* 11/20/2003
/*
/* perform some final touches to the farmland dataset
/*
/*****************************************************************************

&if [exists farmland_alb -cover] &then
  kill farmland_alb  all

copy fm_c_w_f_sc farmland_alb

/* remove unneeded columns carried across
dropitem farmland_alb.pat farmland_alb.pat fm_c_w_f_nsc# fm_c_w_f_nsc-id
farmsc_alb# farmsc_alb-id poly# subclass subclass# rings_ok rings_nok

/* add the final farm id
additem farmland_alb.pat farmland_alb.pat FarmId 8 8 I
additem farmland_alb.pat farmland_alb.pat acres 8 18 f 5


tables
select farmland_alb.pat
calc acres = area / 43560
reselect class = ''
calc class = 'null'
select farmland_alb.pat
reselect class <> 'null'
calc FarmId = $RECNO
quit
```

APPENDIX J. AML SCRIPTS FOR REAPPORTIONING USAGE

```
/* copy and initialize data
/*********************************************
/*&r PrepareInputData.aml - do it manually -

/*********************************************
/* separate the fields from the remaining
/* farmland for the two pilot years
/*********************************************
&r Extract_Field_Specific.aml 95
/*&r Extract_Field_Specific.aml 96
/*&r Extract_Field_Specific.aml 97
/*&r Extract_Field_Specific.aml 98
/*&r Extract_Field_Specific.aml 99
/*&r Extract_Field_Specific.aml 00
/*&r Extract_Field_Specific.aml 01

/*********************************************
/* mtrs fractions
/*********************************************
&r process_mtrs_fractions.aml 95
/*&r process_mtrs_fractions.aml 96
/*&r process_mtrs_fractions.aml 97
/*&r process_mtrs_fractions.aml 98
/*&r process_mtrs_fractions.aml 99
/*&r process_mtrs_fractions.aml 00
/*&r process_mtrs_fractions.aml 01

/*********************************************
/* fields fractions
/*********************************************
&r process_field_fractions.aml 95
/*&r process_field_fractions.aml 96
/*&r process_field_fractions.aml 97
/*&r process_field_fractions.aml 98
/*&r process_field_fractions.aml 99
/*&r process_field_fractions.aml 00
/*&r process_field_fractions.aml 01

&type All Done.
```

DefineProjection.aml

```
/*****************************************************************************
/* Angelita C. - California Schools Project
/* Created by: Matt Werth, ICF Consulting GIS Dept
/* 1/26/2004
/*
/* define projection
/*
/*****************************************************************************
```

```
&args  cov

&if %cov% = "" &then
  &return ERROR - Define projection expected a coverage name

projectdefine cover %cov%
Projection             ALBERS
Datum                   NAD83
Units                  FEET
Spheroid               GRS1980
Parameters
 32  0  0.000
 42  0  0.000
 -120  0  0.00
 30  0  0.000
 0.00000
 0.00000


/*tolerance %cov% fuzzy 0.000001
/*tolerance %cov% snap 0.000001
/*tolerance %cov% nodesnap 0.000001
/*tolerance %cov% edit 0.000001
/*tolerance %cov% weed 0.000001
/*tolerance %cov% grain 0.000001
```

Extract_Field_Specific.aml

```
/*****************************************************************************
/* Angelita C. - California Schools Project
/* Created by: Matt Werth, ICF Consulting GIS Dept
/* 12/22/2003
/*
/* This function queries the fields specific layer (MeBr_Fields) and extracts
/* the polygons for the requested year, saves them into a year specific layer
/* and produces a copy of the master farmland with the fields specific polygons
/* removed.
/*
/*****************************************************************************
/* take the year to process as an argument
&args strYear
&s sYearExt = _%strYear%
&s f = ../../data/farmland/farmland
&s sIn_FieldsCov = ../../data/farmland/mebr_fields

&r DefineProjection.aml %sIn_FieldsCov%

&s sYearColumn = Y%strYear%
&s sOut_FieldCov = fields%sYearExt%
&s sOut_FarmCov = %f%%sYearExt%

&if [exists %sOut_FieldCov% -cover] &then
  kill %sOut_FieldCov% all

/* select the polygons from the specified year into a seperate file
```

```
arcedit
edit %sIn_FieldsCov% poly
select all
resel %sYearColumn% = 1
put %sOut_FieldCov%
quit

&r DefineProjection.aml %sOut_FieldCov%

build  %sOut_FieldCov%

&if [exists %sOut_FarmCov% -cover] &then
  kill %sOut_FarmCov% all

/* erase the fields from the general farmland like a cookie cutter
erase %f% %sOut_FieldCov% %sOut_FarmCov%

&r DefineProjection.aml %sOut_FarmCov%

build %sOut_FarmCov%
```

Process_Field_Fractions.aml

```
/*****************************************************************************
/* Angelita C. - California Schools Project
/* Created by: Matt Werth, ICF Consulting GIS Dept
/* 12/22/2003
/*
/* This process will generate a fraction file for a specified year
/*
/*   *** THIS ASSUMES THAT INPUT LAYERS HAVE BEEN PREPARED
/*****************************************************************************

precision double double

/* take the year to process as an argument
&args sYear

&s g = ../../data/modelgrid/mgrid_poly
&s f = fields_%sYear%
&s g_f = g_fields_%sYear%
&s sPath = P:\CASchools\Processing\FractionsState\

/*******************************************************************************************
/* define projection
/*******************************************************************************************
&type ================================================
&type define projection...
&r DefineProjection.aml %g%
&r DefineProjection.aml %f%


/*******************************************************************************************
/* Build Polygons so that the identity operation properly tag items
/*******************************************************************************************
```

```
&type ================================================
&type Building polygons...

&describe %f%
&if %dsc$qedit% &then
  build %f%
&describe %g%
&if %dsc$qedit% &then
  build %g%

/*****************************************************************************
/* additem area_f to the farmland polygon so we can remember the original farm field area
/* before the intersection takes place.
/*****************************************************************************
&type ================================================
&type add AREA_F to farm fields...

&if [iteminfo %f%.pat -info area_f -exists] &then
dropitem %f%.pat %f%.pat area_f

additem %f%.pat %f%.pat area_f 8 18 f 5

tables
select %f%.pat
calc area_f = area
quit

/*****************************************************************************
/* Overlay Grid with farmland
/*****************************************************************************
&type ================================================
&type Overlay Fields with Grid...

&if [exists %g_f% -cover] &then
  kill %g_f% all

identity %g% %f% %g_f% poly 0.00001


/*****************************************************************************
/* define projection
/*****************************************************************************
&type ================================================
&type define projection...
&r DefineProjection.aml %g_f%

/*****************************************************************************
/* Add and populate AREA_GF to remember area of individual pieces of farmland
/*****************************************************************************
&type ================================================
&type add AREA_GF...

additem %g_f%.pat %g_f%.pat area_gf 8 18 f 5

tables
select %g_f%.pat
```

```
reselect ICF_ID > 0
calc area_gf = area
quit

/*****************************************************************************
/* create and populate the RATIO column
/*****************************************************************************
&type ================================================
&type *** create and populate MULT column...

additem %g_f%.pat %g_f%.pat Mult 8 18 f 10

tables
select %g_f%.pat
calc Mult = 0
reselect ICF_ID > 0
calc Mult = area_gf / area_f
quit


/*****************************************************************************
/* Finally, create the output fraction file
/*****************************************************************************
&type ================================================
&type creating clean info file and exporting

&if [exists Fractions.dat -info] &then
  killINFO Fractions.dat

pullitems %g_f%.pat fractions.dat %g_f%-ID GRIDID, ICF_ID, MTRS, Area_gf, area_f, Mult

/* create/overwrite text file with containing column headers
&SYS echo ROW-ID, GRIDID, ICF_ID, MTRS, Area_gf, area_f, Mult >
%sPath%fractions_Fields_%sYear%.txt

&Data Arc Info
          ARC
          SELECT FRACTIONS.DAT
     RESELECT ICF_ID > 0
          EXPORT %sPath%fractions_Fields_%sYear%.txt SDF APPEND
          Q STOP
&end


/*=====================================================================
/* clean up intermediate garbage files
/*=====================================================================
&type ================================================
&type cleaning up intermediate files....
&type ================================================
&if [exists fractions.dat -info] &then
   killinfo fractions.dat


&type ================================================
&type Field Fractions are complete.  Have a nice day.
```

```
&type ================================================
```

Process_MTRS_Fractions.aml

```
/*************************************************************************
/* Angelita C. - California Schools Project
/* Created by: Matt Werth, ICF Consulting GIS Dept
/* 12/22/2003
/*
/* This process will generate a fraction file for a specified year
/*
/*   *** THIS ASSUMES THAT INPUT LAYERS HAVE BEEN PREPARED
/*************************************************************************

/* take the year to process as an argument
&args sYear

&s g = ../../data/modelgrid/mgrid_poly
&s m = ../../data/MTRS/mtrs_usage
&s f = farmland_%sYear%

&s m_f = m_f_%sYear%
&s m_f_g  = m_f_g_%sYear%
&s sPath = P:\CASchools\Processing\FractionsState\
&s dblPseudoThreshold_PercentFarmland 0.02

/*****************************************************************************
/* define projection
/*****************************************************************************
&type ================================================
&type define projection...
&r DefineProjection.aml %g%
&r DefineProjection.aml %m%
&r DefineProjection.aml %f%

/*****************************************************************************
/* Build Polygons so that the identity operation properly tag items
/*****************************************************************************
&s time1 = [date -full]

&describe %m%
&if %dsc$qedit% &then
  build %m%
&describe %f%
&if %dsc$qedit% &then
  build %f%
&describe %g%
&if %dsc$qedit% &then
  build %g%


/*****************************************************************************
/* Overlay MTRS with farmland
/*****************************************************************************
&type ================================================
```

```
&type Overlay MTRS with Farmland

&if [exists %m_f% -cover] &then
  kill %m_f% all

identity %m% %f% %m_f% poly 0.00001

&r DefineProjection.aml %m_f%

/*************************************************************
/* overlay m_f with model g
/*****************************************************************************************
&type ================================================
&type Overlaying M_F_G

&if [exists %m_f_g% -cover] &then
  kill %m_f_g% all

identity %m_f% %g% %m_f_g% poly 0.00001

&r DefineProjection.aml %m_f_g%

/*****************************************************************************************
/* add item area_gf
/*****************************************************************************************
&type ================================================
&type add AREA_GF

&if [iteminfo %m_f_g%.pat -info area_gf -exists] &then
          dropitem %m_f_g%.pat %m_f_g%.pat area_gf
additem %m_f_g%.pat %m_f_g%.pat area_gf 8 18 f 5


/*****************************************************************************************
/* select the farmland polygons and populate item area_gf
/*****************************************************************************************
&type ================================================
&type *** Populate  area_GF

tables
select %m_f_g%.pat
calc area_gf = 0
reselect farmid <> 0
calc area_gf = area
quit


/*****************************************************************************************
/* sum the area of farmland inside each MTRS as area_sumf.
/*****************************************************************************************
&type ================================================
&type *** Sum Farmland Area Within MTRS

/* remove DAT file
&if [exists %m_f_g%.dat -info] &then
   killinfo %m_f_g%.dat
```

```
/* sum area of each piece of farmland totaling by MTRS
frequency %m_f_g%.pat %m_f_g%.dat
mtrsid
end
area_gf
end

/* copy the result column into the PAT file
additem %m_f_g%.dat %m_f_g%.dat area_sumf 8 18 f 5
tables
select %m_f_g%.dat
calc area_sumf = area_gf
quit

/* drop the temp column so it doesn't carry along into the final
dropitem %m_f_g%.dat %m_f_g%.dat area_gf

/* add an index so the join doesn't take hours
&type ================================================
&type creating indexes on MTRS
INDEXITEM %m_f_g%.pat MTRSID
INDEXITEM %m_f_g%.dat MTRSID

&if [iteminfo %m_f_g%.pat -info area_sumf -exists] &then
        dropitem %m_f_g%.pat %m_f_g%.pat area_sumf

/* join the DAT to the PAT to bring the area calc across into the PAT
joinitem %m_f_g%.pat %m_f_g%.dat %m_f_g%.pat mtrsid



/*****************************************************************************
/* Assign Psuedo Farmland
/* In case there is no farmland inside a MTRS, we populate the area_mf with area_m
/*****************************************************************************
&type ================================================
&type *** Identifying and assigning pseudo farmland

&if [iteminfo %m_f_g%.pat -info percent_f -exists] &then
        dropitem %m_f_g%.pat %m_f_g%.pat percent_f
additem %m_f_g%.pat %m_f_g%.pat percent_f 8 18 f 5

tables
select %m_f_g%.pat
calc percent_f = 0
select %m_f_g%.pat
reselect area_m > 0
calc percent_f = area_sumf / area_m
select %m_f_g%.pat
reselect percent_f <= %dblPseudoThreshold_PercentFarmland% and farmid = 0
calc Class = 'Pseudo'
calc area_gf = area
quit
```

```
/*****************************************************************************
/* SUM the Area again now that we have evaluated and included Pseudo farmland
/*****************************************************************************
&type ================================================
&type *** Sum Farmland Area Within MTRS

/* remove DAT file
&if [exists %m_f_g%.dat -info] &then
   killinfo %m_f_g%.dat

/* sum area of each piece of farmland totaling by MTRS
frequency %m_f_g%.pat %m_f_g%.dat
mtrsid
end
area_gf
end

/* copy the result column into the PAT file
additem %m_f_g%.dat %m_f_g%.dat area_sumf 8 18 f 5
tables
select %m_f_g%.dat
calc area_sumf = area_gf
quit

/* drop the temp column so it doesn't carry along into the final
dropitem %m_f_g%.dat %m_f_g%.dat area_gf

/* add an index so the join doesn't take hours
&type ================================================
&type creating indexes on MTRS
INDEXITEM %m_f_g%.dat MTRSID

&if [iteminfo %m_f_g%.pat -info area_sumf -exists] &then
        dropitem %m_f_g%.pat %m_f_g%.pat area_sumf

/* join the DAT to the PAT to bring the area calc across into the PAT
joinitem %m_f_g%.pat %m_f_g%.dat %m_f_g%.pat mtrsid



/*****************************************************************************
/* create and populate the MULT column
/*****************************************************************************
&type ================================================
&type *** create and populate MULT column

&if [iteminfo %m_f_g%.pat -info Mult -exists] &then
        dropitem %m_f_g%.pat %m_f_g%.pat Mult
additem %m_f_g%.pat %m_f_g%.pat Mult 8 18 f 10


tables
select %m_f_g%.pat
calc Mult = 0
select %m_f_g%.pat
reselect area_sumf > 0 and area_gf > 0
```

```
calc Mult = area_gf / area_sumf
quit


/*****************************************************************************
/* Finally, create the output fraction file
/*****************************************************************************
&type ================================================
&type *** creating clean info file and exporting
&if [exists fractions.dat -info] &then
  killINFO fractions.dat

pullitems %m_f_g%.pat fractions.dat %m_f_g%-ID MTRS GridId FarmId Class Area_sumF Area_GF Mult

/* create/overwrite text file with containing column headers
&SYS echo ROW-ID, MTRS, GridId, FarmId, Class, Area_sumf, Area_GF, Mult >
%sPath%fractions_MTRS_%sYear%.txt

&Data Arc Info
          ARC
          SELECT FRACTIONS.DAT
     RESELECT MTRS <> ''
          EXPORT %sPath%fractions_MTRS_%sYear%.txt SDF APPEND
          Q STOP
&end


/*****************************************************************************
/* All done
/*****************************************************************************
&s endtime = [date -full]
&type All Done.  Have a swell day! time:%endtime%
```

AUXILIARY AMLS

MTRS_Area.aml

```
/*****************************************************************************
/* additem area_m to the polygon attribute table
/* notice we need to sum the area to account for multiple polygons
/* with the same MTRS section ID
/*****************************************************************************
&s m = mtrs_pi_usg
&if [iteminfo %m%.pat -info tmp_area -exists] &then
dropitem %m%.pat %m%.pat tmp_area
additem %m%.pat %m%.pat tmp_area 8 12 f 3
tables
select %m%.pat
calc tmp_area = area
quit
/* remove DAT file
&if [exists %m%.dat -info] &then
   killinfo %m%.dat
/* sum area of each piece of farmland totaling by MTRS
frequency %m%.pat %m%.dat
mtrsid
end
tmp_area
end
/* copy the result column into the PAT file
additem %m%.dat %m%.dat area_m 8 12 f 3
tables
select %m%.dat
calc area_m = tmp_area
select %m%.dat
resel MTRSID = 0
purge
y
resel MTRSID = 999999999
purge
y
save
quit

/* drop the temp column so it doesn't carry along into the final
dropitem %m%.dat %m%.dat tmp_area
/* join the DAT to the PAT to bring the area calc across into the PAT
joinitem %m%.pat %m%.dat %m%.pat mtrsid
&if [exists %m%.dat -info] &then
   killinfo %m%.dat
```

PrepareInputData.aml

```
/*****************************************************************************
/* Angelita C. - California Schools Project
/* Created by: Matt Werth, ICF Consulting GIS Dept
/* 12/22/2003
/*****************************************************************************
&s g = ../Model_Grid/g
&s InputMTRS = ../MTRS/mtrs_120403
```

```
&s m = m
&s f = f
/*****************************************************************************
/* additem area_m to the polygon attribute table
/*****************************************************************************
&type copy MTRS
&if [exists %m% -cover] &then
   kill %m% all
copy %InputMTRS% %m%
&r MTRS_Area.aml

/*****************************************************************************
/* copy the master farmland and select only the positive farmland
/*****************************************************************************
&if [exists %f% -cover] &then
   kill %f% all
arcedit
edit ../Farmland/farmland_alb poly
select class = 'FarmDoc' or class = 'FarmDow' or class = 'FarmFrap' or  class = 'Santa Cruz'
put %f%
/*save
quit
build %f%


/*****************************************************************************
/* Build Polygons so that the identity operation properly tag items
/*****************************************************************************
&s time1 = [date -full]

&describe %m%
&if %dsc$qedit% &then
  build %m%
&describe %f%
&if %dsc$qedit% &then
  build %f%
&describe %g%
&if %dsc$qedit% &then
  build %g%

&type ================================================
&type All Done.  Have a nice day
```

APPENDIX K. SAMPLE FRACTION TABLE

RowID – unique ID for each fraction calculation
MTRS – Meridian Township Range Section Identifier
GridID - unique ID for the ¼ mile model gridcell for which the fraction of the overlaying MTRS sections are being calculated
FarmID – unique farmland polygon ID
Class - farmland data source {FarmDoc = CA Department of Conservation, FMMP; Pseudo = no farmland present therefore the entire MTRS is considered to be farmland for modeling purposes since usage has been reported; FarmDow = CA Department of Water Resources, LUSD; Santa Cruz = Santa Cruz County, CA farm data; FarmFRAP = CA Department of Forestry and Fire Protection, FRAP; and, blank where no farmland or reported usage is present}[46]
Area_sumf – area of the entire overlaying farmland polygon within the MTRS section in question
Area_GF – area of the spatial intersection of a single model gridcell and farmland
Mult – the fraction of the overlaying MTRS section that will determine the fraction of MeBr contributed from the MTRS to the model gridcell (calculated by dividing Area_GF by Area_sumF)

ROW-ID, MTRS, GridId, FarmId, Class, Area_sumf, Area_GF, Mult
7484,"M12S02E20",1266246,11919,"FarmDoc",18622271.27945,186061.49877,0.009991
7485,"M12S02E20",1266247,11919,"FarmDoc",18622271.27945,327410.51500,0.017581
7486,"M12S02E20",1266248,11919,"FarmDoc",18622271.27945,325216.11946,0.017463
7487,"M12S02E20",1266248,0," ",18622271.27945,0.00000,0.000000
7488,"M12S02E20",1266249,0," ",18622271.27945,0.00000,0.000000
7489,"M12S02E21",1266249,0," ",3856447.47586,0.00000,0.000000
7490,"M12S02E21",1266250,0," ",3856447.47586,0.00000,0.000000
7491,"M12S02E21",1266250,11966,"FarmDoc",3856447.47586,142513.01285,0.036954
7492,"M12S02E21",1266250,0," ",3856447.47586,0.00000,0.000000
7493,"M12S02E21",1266251,0," ",3856447.47586,0.00000,0.000000
7494,"M12S02E21",1266252,0," ",3856447.47586,0.00000,0.000000
7495,"M12S02E22",1266252,0,"Pseudo",27282628.00391,78194.14494,0.002866
7496,"M12S02E22",1266253,0,"Pseudo",27282628.00391,408024.56703,0.014955
7497,"M12S02E22",1266254,0,"Pseudo",27282628.00391,176406.57337,0.006465
7498,"M12S02E22",1266254,11970,"FarmDoc",27282628.00391,189699.01162,0.006953

[46] See farmland documentation for fuller explanation of these farmland polygon sources.

APPENDIX L. PARAMETERS OF DIGITIZED FIELD MAPS

PERMIT	SITEID	ACRES	NAME/LOCATION	MTRS
27P029A	010001	102.1	Porter Ranch	M12S02E09
27P029A	040001	130.9	McGowan	M12S02E16
27P032A	070001	91.8	Wilder	M12S02E16
27P032A	010001	30.5	Aromas	M12S03E17
27P032A	040001	29.9	Eiskamp	M12S02E16
27P032A	050001	68.9	Sakata Pajaro	M12S02E10
27P032A	020001	35.9	Milladan	M12S02E10
27P032A	080001	21.7	Marinovitch	M12S02E10
27P032A	090001	35.7	Madesko	M12S02E10
27P032A	030001	91.9	Connell	M12S02E12
27S075A	020001	179.8	Dayton	M12S04E30
27S075A	040001	59.1	Silacci	M14S04E30
27S075A	030001	115.0	Bardin	M14S03E24
27P092A	060001	64.1	Poloni-Cox	M12S02E20
27P092A	070001	46.9	Willoughby	M12S02E19
27P122A	030001	9.8	Bluff Rd	M12S02E31
27P122A	020001	13.9	178 Jensen Rd	M12S02E30
27P122A	040001	31.0	Muzzi Ranch	M12S02E31
27P122A	010001	10.1	Salinas Rd	M12S02E20
27P156A	010001	27.9	1800 San Juan Road	M12S03E18
27P195A	020001	12.3	Springfield Ranch	M13S02E05
27P195A	010001	224.8	Mayou Ranch	M12S02E32
27P195A	060001	34.6	Little Ranch	M12S02E30
27PGN04	2	0.2	Brothers Greenhouse	M12S02E23
27PGN04	1	190.7	Brothers Floral Farm	M12S02E23
27PGN16	multi	24.8	Pajaro Valley Greenhouse	M12S02E15
27S124A	150001	195.3	DSA Ranch	M12S03E18
27S124A	140001	137.9	Curtis Ranch	M12S03E18
27S124A	070001	256.8	Trafton	M12S02E19
27S124A	050001	17.9	Poloni-Cox	M12S02E20
27S268A	010001	105.1	Dedampierre Farms	M14S03E24
440447A	010001	17.0	FM Brothers	M12S02E23
440655A	11B	38.2	Gargiulo	M11S02E33
440655A	14A	108.1	Gargiulo	M12SO2E09
440232A	3B	14.1	Kimura	
27P001A	040001	13.5	Corralitos	M12S02E29
27P003A	130001	71.6	Trafton	M12S02E16
27P003A	140001	46.1	Hudson	M12S02E16
27P003A	050001	62.2	Wesco Ranch	M12S02E16
27P038A	020001	20.3	Howard Tao	M12S02E15
27P045A	030001	72.4	McGowan (Yamaoka)	

PERMIT	SITEID	ACRES	NAME/LOCATION	MTRS
27P038A	040001	10.1	Shutters Ranch	M12S02E10
27P049A	020001	63.2	Sheehy	M12S02E11
27S217A	050001	34.5	Ramerez Brothers	M15S04E05
27S217A	050001	21.9	Ramerez Brothers	M15S04E05
56X0048	01	152.5	Rio Mesa Farms - Site 1	
56X0048	02	120.4	Rio Mesa Farms - Site 2	
56C0618	02	118.2	Las Posas Berry Farms	
56X0048	04	101.9	Rio Mesa Farms - Site 4	
56X0026	05	121.4	Saticoy Farm - Site 5	
56X0054	05	45.0	Kotake Farms	
56X0026	08	93.7	Saticoy Farm - Site 8	
56C0676	01	207.2	Santa Rosa Farms	
27P003A	080001	45.1	Wiley Ranch	
440169A	020001	51.1	Fagundes Ranch	
440655A	030001	99.5	Porter Ranch	
27P111A	010001	12.4	Azebedo Ranch	
27S021A	020001	228.9	Jacks Ranch	
27S075A	05006C	9.6	Duncan Ranch	
27S183A	040001	28.0	Cunha Ranch #4	
27X0288	02	153.3	Coastal Berry Farms	
27X0288	03	73.6	Coastal Berry Farms	
27X0288	04	70.1	Coastal Berry Farms	
27X0288	05	73.9	Coastal Berry Farms	
56X0007	05	141.9	Terry Farms	
56X0250	01	215.5	D.W. Berry Farms	
56C0609	01	45.1	T&R Berry Farms	
56C0609	02	42.8	T&R Berry Farms	
56X0209	01	47.1	Conroy's Farms	
56X0276	11	57.2	Solimar Farms	
56C0345	03	10.1	Ramon Hernandez	
56C0577	02	13.5	Blooming Flower Farm	
56X0260	03	37.6	Pacifico - Arimura Ranch	
56X0260	04	100.2	Pacifico - Rose Ranch	
56X0123	01	16.4	Harashima Berry Farms	
56X0171	04	50.6	Salvador Magallon	
56C0566	01	156.4	Mac Berry Farms	
56C0009	03	55.3	C&E Farms	
56C0567	03	74.3	Mariz Berry Farm	
56X0011	02	41.6	Fukutomi Farms	
56X0091	01	58.5	Oxnard Harvest	
56X0114	01	30.4	Nose Brothers Ranch	
56X0055	03	5.0	Iwamoto & Gean Farms	
56X0055	04	8.3	Iwamoto & Gean Farms	
56X0271	02	21.6	Carlos Romero	
56C0475	01	2.2	A.M. Flowers	
56C0029	22	44.9	Somis Pacific Ag	
56X0031	02	29.4	Sumine Rance	

PERMIT	SITEID	ACRES	NAME/LOCATION	MTRS
56X0286	01	14.8	Manuel Robles	
56X0104	01	26.1	Nishi Tatsuo	
56X0209	03	74.9	Conroy's Farm	
56C0566	05	44.5	Mac Berry Farms	
56X0260	05	73.3	Pacifico Berry Farms	
56C0559	05	40.8	Rancho Rosa	
56X0195	15	53.7	Emerald Mist Berry Farms	
56X0195	17	29.0	Emerald Mist Berry Farms	
56X0054	02	24.8	Kotake Farms	
56C0346	03	19.8	Paseo Flowers	
194520A	4-2-1	15.9	Cerritos College	
195054A	5-1-1	39.5	Zamperini Field	
27S094A	010001	356.7	V+L Farms	
27S021A	010001	39.2	Hansen Ranch	
27S001A	04001	91.8	Smith Ranch	
27S258A	02	358.0	Hacienda Farm	
27S075A	020001	43.9	Gabilan View Farm	
27S021A	030001	59.0	B&E Berry	
27S001A	070001	57.5	Madalora Ranch	
195216A	5-1-1	83.6	Moraga Vineyards	
56X0026	01	204.1	Saticoy Berry Farm	
270369A	1A	23.8	Santa Cruz Co	M12S03E05
27P003A	1A	109.1	Santa Cruz Co	M12S02E17
27P006A	1A	20.2	Santa Cruz Co	M11S02E10
27P029A	1A	51.1	Santa Cruz Co	M11S02E36
27P032A	1A	66.6	Santa Cruz Co	M12S03E07
27P032A	3A	83.8	Santa Cruz Co	M12S03E07
27P032A	4A	34.6	Santa Cruz Co	M12S03E07
27P032A	5A	48.2	Santa Cruz Co	M12S02E18
27P032A	6A	46.8	Santa Cruz Co	M12S02E18
27P039A	1A	33.4	Santa Cruz Co	M12S03E05
27P042A	110001	12.9	Santa Cruz Co	M12S03E09
27P048A	1A	5.1	Santa Cruz Co	M11S01E24
27P063A	1A	49.6	Santa Cruz Co	M11S01E25
27P092A	1A	8.1	Santa Cruz Co	M12S01E12
27P106A	1A	12.4	Santa Cruz Co	M12S03E09
27P124A	1A	29.4	Santa Cruz Co	M11S02E28
27P125A	1A	10.4	Santa Cruz Co	M11S03E31
27P155A	1A	54.8	Santa Cruz Co	M12S01E14
27P192A	10A	4.2	Santa Cruz Co	M12S01E02
27P192A	2A	80.4	Santa Cruz Co	M12S01E12
27P192A	4A	50.4	Santa Cruz Co	M12S01E02
27P192A	5A	49.8	Santa Cruz Co	M12S01E02
27S0179	1A	14.4	Santa Cruz Co	M11S01E36
27S0179	2A	17.1	Santa Cruz Co	M11S01E36
27S020A	1A	38.2	Santa Cruz Co	M11S02E15
27S020A	2A	40.4	Santa Cruz Co	M12S03E05

PERMIT	SITEID	ACRES	NAME/LOCATION	MTRS
27S027A	1A	13.3	Santa Cruz Co	M11S02E21
27S117A	1A	5.7	Santa Cruz Co	M12S03E09
27S338A	1A	23.0	Santa Cruz Co	M12S01E01
440004A	3A	30.4	Santa Cruz Co	M11S02E08
440043A	7A	10.3	Santa Cruz Co	M11S02E17
440044A	1A	92.5	Santa Cruz Co	M11S02E33
440064A	2A	8.5	Santa Cruz Co	M11S02E29
440066A	2A	19.1	Santa Cruz Co	M11S02E36
440066A	3A	31.0	Santa Cruz Co	M11S02E26
440066A	4A	21.7	Santa Cruz Co	M11S02E36
440066A	5A	7.7	Santa Cruz Co	M11S02E35
440066A	6A	66.6	Santa Cruz Co	M11S02E35
440066A	7A	47.3	Santa Cruz Co	M12S02E02
440066A	8A	63.1	Santa Cruz Co	M12S02E02
440068A	1A	76.2	Santa Cruz Co	M11S03E31
440068A	2A	53.2	Santa Cruz Co	M11S03E31
440068A	4A	63.9	Santa Cruz Co	M11S03E31
440068A	5A	31.2	Santa Cruz Co	M12S01E25
440068A	6A	2.8	Santa Cruz Co	M12S01E25
440068A	7A	44.8	Santa Cruz Co	M12S03E06
440071A	6A	39.7	Santa Cruz Co	M12S01E10
440071A	9A	32.5	Santa Cruz Co	M11S01E34
440079A	1A	30.2	Santa Cruz Co	M12S02E08
440079A	2A	13.5	Santa Cruz Co	M11S02E36
440079A	3A	6.8	Santa Cruz Co	M12S02E18
440079A	4A	8.2	Santa Cruz Co	M12S02E18
440079A	6A	14.7	Santa Cruz Co	M12S02E03
440080A	2A	16.7	Santa Cruz Co	M11S02E17
440080A	3A	7.0	Santa Cruz Co	M10S03W09
440081A	11A	26.2	Santa Cruz Co	M11S03E31
440138A	3A	32.7	Santa Cruz Co	M12S01E13
440138A	4A	18.8	Santa Cruz Co	M12S01E02
440138A	5A	32.5	Santa Cruz Co	M11S02E16
440138A	6A	23.0	Santa Cruz Co	M11S02E16
440141A	1A	20.6	Santa Cruz Co	M11S02E19
440147A	12A	45.3	Santa Cruz Co	M11S02E27
440147A	2A	15.3	Santa Cruz Co	M11S02E26
440147A	3A	27.8	Santa Cruz Co	M11S02E26
440147A	4A	44.5	Santa Cruz Co	M11S02E26
440147A	6A	37.6	Santa Cruz Co	M11S02E26
440147A	9A	27.5	Santa Cruz Co	M11S02E26
440148A	1B	7.5	Santa Cruz Co	M12S01E25
440148A	4A	18.9	Santa Cruz Co	M12S01E25
440164A	1A	110.6	Santa Cruz Co	M11S02E19
440167A	1A	7.9	Santa Cruz Co	M11S01E24
440167A	2A	3.9	Santa Cruz Co	M11S01E24
440168A	1A	96.0	Santa Cruz Co	M11S02E33

PERMIT	SITEID	ACRES	NAME/LOCATION	MTRS
440168A	3A	61.1	Santa Cruz Co	M11S02E35
440169A	14A	46.1	Santa Cruz Co	M11S02E28
440169A	16A	22.4	Santa Cruz Co	M11S01E24
440169A	2A	38.3	Santa Cruz Co	M11S02E25
440169A	3A	93.8	Santa Cruz Co	M11S02E33
440169A	4A	6.6	Santa Cruz Co	M11S02E29
440172A	1A	116.1	Santa Cruz Co	M12S02E06
440175A	7A	8.4	Santa Cruz Co	M11S02E26
440198A	1A	17.5	Santa Cruz Co	M11S02E21
440202A	1A	32.1	Santa Cruz Co	M11S03E31
440202A	3A	9.7	Santa Cruz Co	M11S03E31
440202A	4A	20.1	Santa Cruz Co	M11S03E31
440202A	9A	78.2	Santa Cruz Co	M11S03E31
440211A	1A	168.0	Santa Cruz Co	M12S01E13
440224A	1A	20.3	Santa Cruz Co	M12S01E12
440224A	2A	24.6	Santa Cruz Co	M12S01E13
440230A	1A	136.4	Santa Cruz Co	M11S02E26
440232A	1A	7.8	Santa Cruz Co	M11S02E34
440232A	2A	9.5	Santa Cruz Co	M11S02E34
440237A	1A	24.3	Santa Cruz Co	M12S02E03
440243A	1A	67.4	Santa Cruz Co	M12S01E02
440243A	2A	69.3	Santa Cruz Co	M12S01E11
440243A	3A	82.5	Santa Cruz Co	M12S01E11
440245A	1A	23.3	Santa Cruz Co	M12S01E13
440245A	3A	138.3	Santa Cruz Co	M12S01E13
440254A	1A	8.9	Santa Cruz Co	M12S01E14
440261A	10A	23.0	Santa Cruz Co	M12S02E02
440261A	1A	28.7	Santa Cruz Co	M11S02E35
440261A	4A	33.2	Santa Cruz Co	M12S03E06
440261A	9A	26.5	Santa Cruz Co	M12S02E03
440270A	1A	117.6	Santa Cruz Co	M11S02E36
440270A	2A	86.4	Santa Cruz Co	M12S02E01
440280A	1A	17.0	Santa Cruz Co	M11S01E03
440306A	1A	20.4	Santa Cruz Co	M12S01E02
440306A	3A	14.5	Santa Cruz Co	M11S02E22
440319A	11A	61.9	Santa Cruz Co	M11S02E15
440319A	13A	11.3	Santa Cruz Co	M11S02E15
440319A	2A	53.8	Santa Cruz Co	M11S02E25
440319A	7A	12.8	Santa Cruz Co	M11S02E15
440319A	8A	49.6	Santa Cruz Co	M11S02E15
440323A	1A	5.3	Santa Cruz Co	M11S01E25
440344A	2A	57.1	Santa Cruz Co	M12S02E18
440357A	1A	64.6	Santa Cruz Co	M11S02E30
440357A	5A	36.5	Santa Cruz Co	M11S02E30
440367A	1A	5.7	Santa Cruz Co	M12S01E01
440368A	1A	11.5	Santa Cruz Co	M12S03E06
440368A	2A	20.6	Santa Cruz Co	M11S02E19

PERMIT	SITEID	ACRES	NAME/LOCATION	MTRS
440371A	2A	13.3	Santa Cruz Co	M11S01E36
440371A	3A	29.2	Santa Cruz Co	M12S01E01
440377A	1A	17.9	Santa Cruz Co	M12S01E13
440385A	2A	21.5	Santa Cruz Co	M11S01E27
440394A	3A	22.7	Santa Cruz Co	M12S01E13
440394A	4A	21.0	Santa Cruz Co	M12S02E08
440394A	6A	9.5	Santa Cruz Co	M12S02E08
440413A	1A	62.2	Santa Cruz Co	M12S01E11
440454A	1A	27.1	Santa Cruz Co	M11S02E26
440461A	1A	30.6	Santa Cruz Co	M11S02E35
440478A	1A	85.7	Santa Cruz Co	M12S01E14
440502A	1A	30.1	Santa Cruz Co	M11S02E36
440585A	11A	19.9	Santa Cruz Co	M11S02E19
440585A	3A	10.4	Santa Cruz Co	M11S02E28
440585A	4A	23.5	Santa Cruz Co	M11S02E28
440585A	8A	38.1	Santa Cruz Co	M11S02E28
440651A	1A	90.3	Santa Cruz Co	M11S02E16
440655A	15A	32.4	Santa Cruz Co	M12S02E17
440655A	4A	96.2	Santa Cruz Co	M12S02E17
440669A	3A	28.3	Santa Cruz Co	M12S02E03
440669A	4A	26.6	Santa Cruz Co	M12S02E03
440669A	5A	12.7	Santa Cruz Co	M12S02E03
440669A	6A	15.7	Santa Cruz Co	M11S02E28
440669A	7A	15.1	Santa Cruz Co	M11S02E28
440669A	8A	13.7	Santa Cruz Co	M11S02E28
440669A	9A	15.6	Santa Cruz Co	M11S02E28
440672A	1A	37.3	Santa Cruz Co	M12S01E03
440672A	2A	22.9	Santa Cruz Co	M12S01E03
441047A	2A	12.7	Santa Cruz Co	M11S01E13
441067A	4B	56.1	Santa Cruz Co	M11S02E17
441269A	1A	3.3	Santa Cruz Co	M11S01E34
441282A	1A	8.0	Santa Cruz Co	M12S01E03
443000A	1A	36.3	Santa Cruz Co	M12S01E12
443000A	2A	37.3	Santa Cruz Co	M12S01E11

APPENDIX M. COUNTY CONTACTS AT THE AGRICULTURAL COMMISSIONER OFFICES

Santa Cruz
Lisa LeCoump
Santa Cruz Agricultural Commissioner Office
(831) 763-8080
agc024@park.co.santa-cruz.ca.us

John Kenney
agc033@park.co.santa-cruz.ca.us

Monterey
Ken E. Allen
Monterey County Agricultural Office
Salinas, CA
(831) 759-7307 (direct number)
allenke@co.monterey.ca.us

Ventura
Susan Johnson
Deputy Agricultural Commissioner
Susan.Johnson@mail.co.ventura.ca.us

Kern
David Moore
Deputy Agricultural Commissioner/Sealer
(661) 868-6300
mooreda@co.kern.ca.us

APPENDIX N. ARCOBJECTS SCRIPT FOR QUALITY ASSURANCE OF AREA CALCULATIONS

AssemblyInfo.vb

```
Imports System.Reflection
Imports System.Runtime.InteropServices

' General Information about an assembly is controlled through the following
' set of attributes. Change these attribute values to modify the information
' associated with an assembly.

' Review the values of the assembly attributes

<Assembly: AssemblyTitle("")>
<Assembly: AssemblyDescription("")>
<Assembly: AssemblyCompany("")>
<Assembly: AssemblyProduct("")>
<Assembly: AssemblyCopyright("")>
<Assembly: AssemblyTrademark("")>
<Assembly: CLSCompliant(True)>

'The following GUID is for the ID of the typelib if this project is exposed to COM
<Assembly: Guid("FBCDE468-8FDE-48B1-9192-BCCBC0C0B999")>

' Version information for an assembly consists of the following four values:
'
'      Major Version
'      Minor Version
'      Build Number
'      Revision
'
' You can specify all the values or you can default the Build and Revision Numbers
' by using the '*' as shown below:

<Assembly: AssemblyVersion("1.0.*")>
```

clsAppConfig.vb

```
Imports System.Reflection

Public Class clsAppConfig

    ''*****************************************************************************
    ''*****************************************************************************
    Public Function AppPath() As String
        Try
            Dim strPath As String

            ''Get the app path
            Dim pAssembly As System.Reflection.Assembly
            pAssembly = pAssembly.GetAssembly(Me.GetType())
            strPath = System.IO.Path.GetDirectoryName(pAssembly.Location)

            If Right(strPath, 1) <> "\" Then
                strPath = strPath & "\"
            End If

            '' this doesn't work when this class is in a DLL. -> Executing assemebly is the host app
            ''string AppName = System.Reflection.Assembly.GetExecutingAssembly().FullName
            'Dim strAppFileName As String = System.Reflection.Assembly.GetExecutingAssembly().Location
            'Dim intIndex As Integer = strAppFileName.LastIndexOf("\\")
            'Dim strPath As String = strAppFileName.Substring(0, intIndex + 1)

            Return strPath
```

```
        Catch ex As System.Exception

            MsgBox("Error getting AppPath(). '" & ex.Message & "'")
            Return ""
        End Try
    End Function

    ''*************************************************************************
    ''*************************************************************************
    Public Function GetConfigCollection(ByVal strItem As String) As Collection
        Dim sValue As String
        Dim pRetCollection As Collection = New Collection()
        Try
            Dim sAppPath As String = AppPath()
            Dim sXmlConfigFile As String = sAppPath & "AppConfig.xml"

            Dim pXmlDoc As System.Xml.XmlDocument = New System.Xml.XmlDocument()
            pXmlDoc.Load(sXmlConfigFile)

            Dim pXmlNode As Xml.XmlNode
            pXmlNode = pXmlDoc.SelectSingleNode(strItem)

            Dim iRow As Integer
            For iRow = 0 To pXmlNode.ChildNodes.Count - 1
                sValue = pXmlNode.ChildNodes.Item(iRow).InnerText()
                pRetCollection.Add(sValue)
            Next

            Return pRetCollection
        Catch ex As Xml.XmlException
            MsgBox("XML ERROR READING CONFIG FILE.  error: '& ex.Message & ") '"
        End Try
        pRetCollection = Nothing
        Return Nothing
    End Function

    ''*************************************************************************
    ''*************************************************************************
    Public Function GetConfigItem(ByVal strItem As String) As String

        Dim sValue As String
        Try
            Dim sAppPath As String = AppPath()
            Dim sXmlConfigFile As String = sAppPath & "AppConfig.xml"

            Dim pXmlDoc As System.Xml.XmlDocument = New System.Xml.XmlDocument()
            pXmlDoc.Load(sXmlConfigFile)

            sValue = pXmlDoc.SelectSingleNode(strItem).InnerText

        Catch ex As Xml.XmlException
            MsgBox("XML ERROR READING CONFIG FILE.  error: '& ex.Message & ") '"
        End Try

        Return sValue

    End Function
```

End Class

clsDatabase.vb

```
''*********************************************************************************************
'' ICF General Database Connectivity Module
'' Created by: Matt Werth, ICF Consulting GIS Dept
'' 1/8/04
''
'' Database functions
''
```

```
''*****************************************************************************************
Imports System
Imports System.Data
Imports System.Data.OleDb

Public Class clsDatabase

    ''*************************************************************************
    '' constant path to access db
    Dim pAppConfig As clsAppConfig

    ''*************************************************************************
    ''*************************************************************************
    Public Sub New()
        pAppConfig = New clsAppConfig()

        ''strDatabase = AppPath() + strDatabase
    End Sub

    ''*************************************************************************
    ''*************************************************************************
    Protected Overrides Sub Finalize()

        pAppConfig = Nothing
        MyBase.Finalize()
    End Sub


    ''*************************************************************************
    ''*************************************************************************
    Public Function GetDataset(ByVal strDatabase As String, ByVal sSQL As String, ByVal sDatasetName As String) As DataSet
        Try
            '' get the Application variable
            ''string connectString = Application["APPLICATION_ACCOUNT"].ToString()
            ''string connectString = "Provider=MSDAORAData Source=" + sDSN + "User ID=" + sUID + "Password=" + sPWD + ""
            Dim sConnectString As String = "Provider=Microsoft.Jet.OLEDB.4.0;Data Source=" & strDatabase

            ''Create connection
            Dim pConnection As OleDbConnection = New OleDbConnection(sConnectString)

            ''Create adapter
            Dim pAdapter As OleDbDataAdapter = New OleDbDataAdapter()

            ''Create DataSet
            Dim pDataSet As DataSet = New DataSet()

            ''Run select on adapter
            pAdapter.SelectCommand = New OleDbCommand(sSQL, pConnection)

            ''Fill DataSet
            pAdapter.Fill(pDataSet, sDatasetName)

            Return pDataSet

        Catch ex As System.Data.OleDb.OleDbException

            MsgBox("Error connecting to and querying database. ' " + ex.Message + """)
            Return Nothing

        End Try

    End Function
    ''*************************************************************************
    ''*************************************************************************
    Public Function GetDataset(ByVal pConnection As OleDbConnection, ByVal sSQL As String, ByVal sDatasetName As String) As
DataSet
        Try

            ''Create adapter
            Dim pAdapter As OleDbDataAdapter = New OleDbDataAdapter()
```

```
        ''Create DataSet
        Dim pDataSet As DataSet = New DataSet()

        ''Run select on adapter
        pAdapter.SelectCommand = New OleDbCommand(sSQL, pConnection)

        ''Fill DataSet
        pAdapter.Fill(pDataSet, sDatasetName)

        Return pDataSet

    Catch ex As System.Data.OleDb.OleDbException

        MsgBox("Error connecting to and querying database. ' " + ex.Message + """)
        Return Nothing

    End Try

End Function

''*************************************************************************
''*************************************************************************
Public Function ExecSQL(ByVal strDatabase As String, ByVal strSQL As String) As Integer
    Try

        '' get the Application variable
        ''string connectString = Application["APPLICATION_ACCOUNT"].ToString()
        ''string connectString = "Provider=MSDAORAData Source=" + sDSN + "User ID=" + sUID + "Password=" + sPWD + ""
        Dim sConnectString As String = "Provider=Microsoft.Jet.OLEDB.4.0;Data Source=" & strDatabase

        ''Create connection
        Dim pConnection As OleDbConnection = New OleDbConnection(sConnectString)
        pConnection.Open()

        '' create command
        Dim pCommand As OleDbCommand = New OleDbCommand(strSQL, pConnection)

        ''set connection
        pCommand.Connection = pConnection

        '' execute and grab rows affected
        Dim intRowsAffected As Integer = pCommand.ExecuteNonQuery()

        pCommand.Connection.Close()
        pCommand.Connection.Dispose()
        pCommand.Dispose()

        Return intRowsAffected

    Catch ex As System.Data.OleDb.OleDbException

        MsgBox("Error executing SQL in database. '" + ex.Message + """)
        Return -1
    End Try

End Function

Public Function GetConnection(ByVal strDatabase As String)
    Try
        '' get the Application variable
        ''string connectString = Application["APPLICATION_ACCOUNT"].ToString()
        ''string connectString = "Provider=MSDAORAData Source=" + sDSN + "User ID=" + sUID + "Password=" + sPWD + ""
        Dim sConnectString As String = "Provider=Microsoft.Jet.OLEDB.4.0;Data Source=" & strDatabase

        ''Create connection
        Dim pConnection As OleDbConnection = New OleDbConnection(sConnectString)
        pConnection.Open()

        Return pConnection
```

```
        Catch ex As System.Data.OleDb.OleDbException

            MsgBox("Error executing SQL in database. '" + ex.Message + "'")
            Return Nothing
        End Try

        Return Nothing


    End Function

    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    Public Function ExecSQL(ByVal pConnection As OleDbConnection, ByVal strSQL As String) As Integer
        Try

            '' create command
            Dim pCommand As OleDbCommand = New OleDbCommand(strSQL, pConnection)

            ''set connection
            pCommand.Connection = pConnection

            '' execute and grab rows affected
            Dim intRowsAffected As Integer = pCommand.ExecuteNonQuery()

            pCommand.Dispose()

            Return intRowsAffected

        Catch ex As System.Data.OleDb.OleDbException

            MsgBox("Error executing SQL in database. '" + ex.Message + "'")
            Return -1
        End Try

    End Function

    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    '' delete a table from an access database
    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    Public Function DeleteAccessTable(ByVal strDatabase As String, ByVal strTable As String) As Boolean

        Dim pDbe As DAO.DBEngine = New DAO.DBEngineClass()
        Dim pDB As DAO.Database
        Try

            pDB = pDbe.OpenDatabase(strDatabase)
            'Dim pEnum As IEnumerator = pDB.TableDefs.GetEnumerator()
            'pEnum.Reset()
            'pEnum.Current.

            Dim pTable As DAO.TableDefClass = pDB.TableDefs(strTable)
            If Not pTable Is Nothing Then
                pDB.TableDefs.Delete(strTable)
            End If
        Catch
            ' if the table does not exist, there will be an exception
            ' since there is nothing to delete, we just move on
        Finally
            pDB.Close()
            pDB = Nothing

        End Try

        Return True

    End Function

End Class
```

FractionWorker.vb

```
''*****************************************************************************************************
'' Angelita C. - California Schools Project
'' Created by: Matt Werth, ICF Consulting GIS Dept
'' 1/8/04
''
'' This class will test the FIELDS fraction files procduced for the Angelita C project.
''
''*****************************************************************************************************
Imports System
Imports System.Data
Imports System.Data.OleDb

Public Class FractionWorker
    Private mpDatabase As clsDatabase
    Private pAppConfig As clsAppConfig

    Private msDataPath As String
    Private msQADB As String
    Private msUsageDB As String
    Private msFractionFileRoot As String = "Fractions_Fields_"
    Private msFractionMTRSRoot As String = "Fractions_MTRS_"
    Private msQALogfile As String
    Private mpProgressWnd As progressWnd
    Private mdMinDif As Double
    Private mdMaxDif As Double
    Private mdMinMult As Double
    Private mdMaxMult As Double
    Private mcolYears As Collection

    ' A creatable COM class must have a Public Sub New()
    ' with no parameters, otherwise, the class will not be
    ' registered in the COM registry and cannot be created
    ' via CreateObject.
    Public Sub New()
        LoadConfig()

    End Sub

    ''*************************************************************************
    ''
    ''*************************************************************************
    Public Function LoadConfig()
        mpDatabase = Nothing
        mpDatabase = New clsDatabase()
        pAppConfig = Nothing
        pAppConfig = New clsAppConfig()

        msDataPath = pAppConfig.GetConfigItem("AppConfiguration/DataPath")
        msFractionFileRoot = pAppConfig.GetConfigItem("AppConfiguration/FractionFileRoot")

        msQADB = pAppConfig.GetConfigItem("AppConfiguration/QADB")
        msUsageDB = pAppConfig.GetConfigItem("AppConfiguration/UsageDB")

        msQALogfile = pAppConfig.GetConfigItem("AppConfiguration/OutputLog")

        mdMinDif = CDbl(pAppConfig.GetConfigItem("AppConfiguration/MinDif"))
        mdMaxDif = CDbl(pAppConfig.GetConfigItem("AppConfiguration/MaxDif"))
        mdMinMult = CDbl(pAppConfig.GetConfigItem("AppConfiguration/MinMult"))
        mdMaxMult = CDbl(pAppConfig.GetConfigItem("AppConfiguration/MaxMult"))

        msQALogfile = msQALogfile & Now().Month & "_" & Now().Day & "_" & Now().Year & "__" & Now.Hour & "_" & Now.Minute &
"_" & Now.Second & ".txt"

        mcolYears = pAppConfig.GetConfigCollection("AppConfiguration/ProcessingYears")

    End Function
```

```
'*************************************************************************
' main entry function called from the command OnClick()
'*************************************************************************
Public Function TestFractions()
   mpProgressWnd = New progressWnd()
   mpProgressWnd.Show()

   WriteLogInfo("*******************************************************************************")
   WriteLogInfo("  Angelita C project, for EPA, by ICF consulting")
   WriteLogInfo("  Fraction File QA: " & Now())
   WriteLogInfo("*******************************************************************************")

   Dim strYear As String
   Dim iYear As Integer
   For iYear = 1 To mcolYears.Count
      strYear = mcolYears.Item(iYear)
      TestMTRSFractions(strYear)
      CheckMTRSUsageagainstMTRSFractions(strYear)
      TestFieldFractions(strYear)
   Next

   mpProgressWnd.Hide()
   mpProgressWnd = Nothing
   MsgBox("Done")

End Function
'*************************************************************************
'
'*************************************************************************
Public Function TestMTRSFractions(ByVal strYear As String)

   WriteLogInfo("*****************************************")
   WriteLogInfo("  MTRS Fractions for year " & strYear)
   WriteLogInfo("  " & Now())
   WriteLogInfo("*****************************************")

   ' push fraction file into database
   Upload_MTRS_FractionFile_To_Database(strYear)

   ' create & populate stats table and populate with groupable sums
   Summarize_MTRS_FractionFile(strYear)

   ReportMTRSStats(strYear)

   WriteLogInfo("")

End Function

'*************************************************************************
'
'*************************************************************************
Public Function TestFieldFractions(ByVal strYear As String)

   WriteLogInfo("*****************************************")
   WriteLogInfo("  FIELD Fractions for year " & strYear)
   WriteLogInfo("  " & Now())
   WriteLogInfo("*****************************************")

   ' push fraction file into database
   Upload_FractionFile_To_Database(strYear)

   ' create & populate stats table and populate with groupable sums
   Summarize_FractionFile(strYear)

   ReportFieldsStats(strYear)

   ListUsageWithNoFractions(strYear)

   ListPolygonsWithNoUsage(strYear)
```

```
    WriteLogInfo("")

End Function

''*************************************************************************
'' uploads a single years field fraction file into the QA database
''*************************************************************************
Public Function Upload_FractionFile_To_Database(ByVal strYear As String) As Boolean

    Dim strFractFile As String
    strFractFile = msDataPath & msFractionFileRoot & strYear & ".txt"

    Dim strFractTable As String = msFractionFileRoot & strYear

    '' Delete the existing table (if exists)
    mpDatabase.DeleteAccessTable(msQADB, strFractTable)

    '' recreate empty table from template
    Dim strSQL As String
    strSQL = "SELECT * INTO " & strFractTable & " FROM " & msFractionFileRoot & "template"
    mpDatabase.ExecSQL(msQADB, strSQL)

    '' get the number of rows in this fraction file
    '' this is a nonessential performance hit, but feedback to user is helpful
    Dim lngRows As Long = GetNumberRowsInTextFile(strFractFile)

    Dim pStreamReader As System.IO.StreamReader = New System.IO.StreamReader(strFractFile)
    If pStreamReader Is Nothing Then
        MsgBox("ERROR, Fraction File Not Found: '" & strFractFile & "'")
        Return False
    End If


    '' get the first row of text, since the first row is actually
    '' the column headers, do it twice
    Dim strRow As String
    strRow = pStreamReader.ReadLine() ' col headers
    strRow = pStreamReader.ReadLine() ' first row of data

    Dim strProgressMessage As String
    strProgressMessage = "Upload " & UCase(msFractionFileRoot & strYear & ".txt") & " into database."

    mpProgressWnd.SetStatus(0, lngRows, strProgressMessage)

    Dim iRow As Long

    ''obtain connection that we can use across updates
    Dim pConnection As OleDbConnection = mpDatabase.GetConnection(msQADB)

    '' loop throught text file and insert rows into database table
    While strRow <> ""

        strSQL = "INSERT INTO " & strFractTable & _
                 " (ROW_ID, GRIDID, ICF_ID, MTRS, Area_gf, area_f, Mult) values " & _
                 "(" & strRow & ")"

        mpDatabase.ExecSQL(pConnection, strSQL)

        iRow = iRow + 1
        If (iRow Mod 5) = 0 Then
            mpProgressWnd.SetStatus(iRow, lngRows, strProgressMessage)
        End If

        strRow = pStreamReader.ReadLine()
    End While
    pConnection.Close()
    pConnection = Nothing

    mpProgressWnd.SetStatus(0, 0, "")
```

```
End Function
'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
'' uploads a single years field fraction file into the QA database
'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
Public Function Upload_MTRS_FractionFile_To_Database(ByVal strYear As String) As Boolean

    Dim strFractFile As String
    strFractFile = msDataPath & msFractionMTRSRoot & strYear & ".txt"

    Dim strFractTable As String = msFractionMTRSRoot & strYear

    '' Delete the existing table (if exists)
    mpDatabase.DeleteAccessTable(msQADB, strFractTable)

    '' recreate empty table from template
    Dim strSQL As String
    strSQL = "SELECT * INTO " & strFractTable & " FROM " & msFractionMTRSRoot & "template"
    mpDatabase.ExecSQL(msQADB, strSQL)

    strSQL = "select ROW_ID, MTRS, GRIDID, FARMID, CLASS, AREA_SUMF, AREA_GF, MULT from " & strFractTable

    ' get empty record set from table
    Dim pUploadSet As DataSet
    pUploadSet = mpDatabase.GetDataset(msQADB, strSQL, strFractTable)

    '' get the number of rows in this fraction file
    '' this is a nonessential performance hit, but feedback to user is helpful
    Dim lngRows As Long = GetNumberRowsInTextFile(strFractFile)

    Dim pStreamReader As System.IO.StreamReader = New System.IO.StreamReader(strFractFile)
    If pStreamReader Is Nothing Then
        MsgBox("ERROR, Fraction File Not Found: '" & strFractFile & "'")
        Return False
    End If

    '' get the first row of text, since the first row is actually
    '' the column headers, do it twice
    Dim strRow As String
    strRow = pStreamReader.ReadLine() ' col headers
    strRow = pStreamReader.ReadLine() ' first row of data

    Dim strProgressMessage As String
    strProgressMessage = "Upload " & UCase(msFractionMTRSRoot & strYear & ".txt") & " into database."

    mpProgressWnd.SetStatus(0, lngRows, strProgressMessage)

    Dim iRow As Long

    "obtain connection that we can use across updates
    Dim pConnection As OleDbConnection = mpDatabase.GetConnection(msQADB)

    '' loop throught text file and insert rows into database table
    While strRow <> ""

        strSQL = "INSERT INTO " & strFractTable & _
                 " (ROW_ID, MTRS, GRIDID, FARMID, CLASS, AREA_SUMF, AREA_GF, MULT) values " & _
                 "(" & strRow & ")"

        mpDatabase.ExecSQL(pConnection, strSQL)

        iRow = iRow + 1
        If (iRow Mod 5) = 0 Then
            mpProgressWnd.SetStatus(iRow, lngRows, strProgressMessage)
        End If

        strRow = pStreamReader.ReadLine()
    End While
    pConnection.Close()
    pConnection = Nothing
```

```
    mpProgressWnd.SetStatus(0, 0, "Updateing database")

    mpProgressWnd.SetStatus(0, 0, "")

End Function

'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
'' open a text file, read in and count the rows
'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
Public Function GetNumberRowsInTextFile(ByVal strTextFile As String) As Long

    Dim pStreamReader As System.IO.StreamReader
    pStreamReader = New System.IO.StreamReader(strTextFile)
    If pStreamReader Is Nothing Then
        MsgBox("ERROR, File Not Found: '" & strTextFile & "'")
        Return -1
    End If

    Dim lngCount As Long
    While pStreamReader.ReadLine() <> ""
        lngCount = lngCount + 1
    End While

    Return lngCount

End Function

'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
'' ·        Create table (SUM_MULT, SUM_GF, AREA_F, AREA_DIF, DB_REC_COUNT)(1/2)
'' ·        Populate Sum(MULT), Sum(AREA_GF), AREA_F group by ICF_ID (1/2)
'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
Public Function Summarize_FractionFile(ByVal strYear As String) As Boolean


    Dim strFractTable As String = msFractionFileRoot & strYear
    Dim strQAFractTable As String = "QA_Fields_" & strYear

    '' Delete the existing table (if exists)
    mpDatabase.DeleteAccessTable(msQADB, strQAFractTable)

    '' recreate empty table from template
    Dim strSQL As String
    strSQL = "SELECT * INTO " & strQAFractTable & " FROM QA_Fields_template"
    mpDatabase.ExecSQL(msQADB, strSQL)

    mpProgressWnd.SetStatus(0, 0, "Summarizing Fraction File")

    ''Do a grouping summarizing query on the fraction table putting the
    '' results into a QA table

    strSQL = "INSERT INTO " & strQAFractTable & " " & _
            "SELECT " & _
              "ICF_ID, " & _
              "Area_f, " & _
              "Sum(Mult) AS SumMult, " & _
              "Sum(Area_GF) AS SumArea_Gf, " & _
              "clng((SumArea_Gf * 1000000) - (Area_f * 1000000)) / 1000000 as AreaDif " & _
            "FROM " & strFractTable & " " & _
            "GROUP BY ICF_ID, Area_f"

    mpDatabase.ExecSQL(msQADB, strSQL)

    mpProgressWnd.SetStatus(0, 0, "")
End Function

'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
'' ·        Create table (SUM_MULT, SUM_GF, AREA_F, AREA_DIF, DB_REC_COUNT)(1/2)
'' ·        Populate Sum(MULT), Sum(AREA_GF), AREA_F group by ICF_ID (1/2)
'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
Public Function Summarize_MTRS_FractionFile(ByVal strYear As String) As Boolean
```

```
    Dim strFractTable As String = msFractionMTRSRoot & strYear
    Dim strQAFractTable As String = "QA_MTRS_" & strYear

    " Delete the existing table (if exists)
    mpDatabase.DeleteAccessTable(msQADB, strQAFractTable)

    " recreate empty table from template
    Dim strSQL As String
    strSQL = "SELECT * INTO " & strQAFractTable & " FROM QA_MTRS_template"
    mpDatabase.ExecSQL(msQADB, strSQL)

    mpProgressWnd.SetStatus(0, 0, "Summarizing MTRS Fractions")

    "Do a grouping summarizing query on the fraction table putting the
    " results into a QA table

    strSQL = "INSERT INTO " & strQAFractTable & " " & _
          "SELECT " & _
            "MTRS, " & _
            "Area_sumf, " & _
            "Sum(Area_GF) AS SumArea_Gf, " & _
            "Sum(Mult) AS SumMult, " & _
            "clng((SumArea_gf * 10000000000) - (Area_sumf * 10000000000)) / 10000000000 as AreaDif " & _
          "FROM " & strFractTable & " " & _
          "GROUP BY MTRS, Area_sumf"

    mpDatabase.ExecSQL(msQADB, strSQL)

    mpProgressWnd.SetStatus(0, 0, "")
  End Function


  '*****************************************************************************
  " list the Usage records with no fractions
  '*****************************************************************************
  Function ReportFieldsStats(ByVal strYear As String)

    "
    " collect states
    "
    Dim strSQL As String
    Dim iRow As Integer
    Dim strRow As String
    Dim pDataset As DataSet
    Dim blnFoundOne As Boolean = False

    "
    " collect and write out top 4 MIN(AREADIF)
    "
    mpProgressWnd.SetStatus(0, 0, "Reporting Summary Stats MIN(AREADIF)")
    strSQL = "SELECT ICF_ID, AREA_F, SUMAREA_GF, SUMMULT, AREADIF FROM QA_Fields_" & strYear & " WHERE
AREADIF < " & mdMinDif & " ORDER BY AREADIF"
    pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
    WriteLogInfo("MIN(AreaDif) < " & mdMinDif)
    blnFoundOne = False
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
      If blnFoundOne = False Then
        WriteLogInfo("icf_id, area_f, sumarea_gf, summult, areadif")
      End If
      strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
            pDataset.Tables(0).Rows(iRow)(1) & ", " & _
            pDataset.Tables(0).Rows(iRow)(2) & ", " & _
            pDataset.Tables(0).Rows(iRow)(3) & ", " & _
            pDataset.Tables(0).Rows(iRow)(4) & ")"

      WriteLogInfo(strRow)
      blnFoundOne = True
    Next
```

```
    If blnFoundOne = False Then
       WriteLogInfo("0 outliers found")
    End If
    WriteLogInfo("")
    pDataset = Nothing

    "
    " collect and write out top 4 MAX(AREADIF)
    "
    mpProgressWnd.SetStatus(0, 0, "Reporting Summary Stats MAX(AREADIF)")
    strSQL = "SELECT ICF_ID, AREA_F, SUMAREA_GF, SUMMULT, AREADIF FROM QA_Fields_" & strYear & " WHERE
AREADIF > " & mdMaxDif & " ORDER BY AREADIF DESC"
    pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
    WriteLogInfo("MAX(AreaDif) > " & mdMaxDif)
    blnFoundOne = False
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
       If blnFoundOne = False Then
          WriteLogInfo("icf_id, area_f, sumarea_gf, summult, areadif")
       End If
       strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
             pDataset.Tables(0).Rows(iRow)(1) & ", " & _
             pDataset.Tables(0).Rows(iRow)(2) & ", " & _
             pDataset.Tables(0).Rows(iRow)(3) & ", " & _
             pDataset.Tables(0).Rows(iRow)(4) & ")"

       WriteLogInfo(strRow)
       blnFoundOne = True
    Next
    If blnFoundOne = False Then
       WriteLogInfo("0 outliers found")
    End If
    WriteLogInfo("")
    pDataset = Nothing

    "
    " collect and write out top 4 MIN(AREADIF)
    "
    mpProgressWnd.SetStatus(0, 0, "Reporting Summary Stats MIN(SUMMULT)")
    strSQL = "SELECT ICF_ID, AREA_F, SUMAREA_GF, SUMMULT, AREADIF FROM QA_Fields_" & strYear & " WHERE
SUMMULT < " & mdMinMult & " ORDER BY SUMMULT"
    pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
    WriteLogInfo("MIN(SumMult) < " & mdMinMult)
    blnFoundOne = False
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
       If blnFoundOne = False Then
          WriteLogInfo("icf_id, area_f, sumarea_gf, summult, areadif")
       End If
       strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
             pDataset.Tables(0).Rows(iRow)(1) & ", " & _
             pDataset.Tables(0).Rows(iRow)(2) & ", " & _
             pDataset.Tables(0).Rows(iRow)(3) & ", " & _
             pDataset.Tables(0).Rows(iRow)(4) & ")"

       WriteLogInfo(strRow)
       blnFoundOne = True
    Next
    If blnFoundOne = False Then
       WriteLogInfo("0 outliers found")
    End If
    WriteLogInfo("")
    pDataset = Nothing


    "
    " collect and write out top 4 MIN(AREADIF)
    "
    mpProgressWnd.SetStatus(0, 0, "Reporting Summary Stats MAX(SUMMULT)")
    strSQL = "SELECT ICF_ID, AREA_F, SUMAREA_GF, SUMMULT, AREADIF FROM QA_Fields_" & strYear & " WHERE
SUMMULT > " & mdMaxMult & " ORDER BY SUMMULT DESC"
    pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
```

```
    WriteLogInfo("MAX(SumMult) > " & mdMaxMult)
    blnFoundOne = False
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
      If blnFoundOne = False Then
        WriteLogInfo("icf_id, area_f, sumarea_gf, summult, areadif")
      End If
      strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
          pDataset.Tables(0).Rows(iRow)(1) & ", " & _
          pDataset.Tables(0).Rows(iRow)(2) & ", " & _
          pDataset.Tables(0).Rows(iRow)(3) & ", " & _
          pDataset.Tables(0).Rows(iRow)(4) & ")"

      WriteLogInfo(strRow)
      blnFoundOne = True
    Next
    If blnFoundOne = False Then
      WriteLogInfo("0 outliers found")
    End If
    WriteLogInfo("")
    pDataset = Nothing
    mpProgressWnd.SetStatus(0, 0, "")
  End Function

  '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
  '' list the Usage records with no fractions
  '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
  Function ReportMTRSStats(ByVal strYear As String)

    ''
    '' collect states
    ''
    Dim strSQL As String
    Dim iRow As Integer
    Dim strRow As String
    Dim pDataset As DataSet

    Dim blnFoundOne As Boolean = False

    ''
    '' collect and write out top 4 MIN(AREADIF)
    ''
    mpProgressWnd.SetStatus(0, 0, "Reporting Summary Stats MIN(AREADIF)")
    strSQL = "SELECT MTRS, AREA_SUMF, SUMAREA_GF, SUMMULT, AREADIF FROM QA_MTRS_" & strYear & " WHERE
AREADIF < " & mdMinDif & " ORDER BY AREADIF"
    pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
    WriteLogInfo("MIN(AreaDif) < " & mdMinDif)
    blnFoundOne = False
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
      If blnFoundOne = False Then
        WriteLogInfo("MTRS, area_sumf, sumarea_gf, summult, areadif")
      End If
      strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
          pDataset.Tables(0).Rows(iRow)(1) & ", " & _
          pDataset.Tables(0).Rows(iRow)(2) & ", " & _
          pDataset.Tables(0).Rows(iRow)(3) & ", " & _
          pDataset.Tables(0).Rows(iRow)(4) & ")"

      WriteLogInfo(strRow)
      blnFoundOne = True
    Next
    If blnFoundOne = False Then
      WriteLogInfo("0 outliers found")
    End If
    WriteLogInfo("")
    pDataset = Nothing

    ''
    '' collect and write out top 4 MAX(AREADIF)
    ''
    mpProgressWnd.SetStatus(0, 0, "Reporting Summary Stats MAX(AREADIF)")
```

```
    strSQL = "SELECT MTRS, AREA_SUMF, SUMAREA_GF, SUMMULT, AREADIF FROM QA_MTRS_" & strYear & " WHERE
AREADIF > " & mdMaxDif & " ORDER BY AREADIF DESC"
    pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
    WriteLogInfo("MAX(AreaDif) > " & mdMaxDif)
    blnFoundOne = False
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
      If blnFoundOne = False Then
        WriteLogInfo("MTRS, area_sumf, sumarea_gf, summult, areadif")
      End If
      strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
          pDataset.Tables(0).Rows(iRow)(1) & ", " & _
          pDataset.Tables(0).Rows(iRow)(2) & ", " & _
          pDataset.Tables(0).Rows(iRow)(3) & ", " & _
          pDataset.Tables(0).Rows(iRow)(4) & ")"

      WriteLogInfo(strRow)
      blnFoundOne = True
    Next
    If blnFoundOne = False Then
      WriteLogInfo("0 outliers found")
    End If
    WriteLogInfo("")
    pDataset = Nothing

    "
    " collect and write out top 4 MIN(AREADIF)
    "
    mpProgressWnd.SetStatus(0, 0, "Reporting Summary Stats MIN(SUMMULT)")
    strSQL = "SELECT MTRS, AREA_SUMF, SUMAREA_GF, SUMMULT, AREADIF FROM QA_MTRS_" & strYear & " WHERE
SUMMULT < " & mdMinMult & " ORDER BY SUMMULT"
    pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
    WriteLogInfo("MIN(SumMult) < " & mdMinMult)
    blnFoundOne = False
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
      If blnFoundOne = False Then
        WriteLogInfo("MTRS, area_sumf, sumarea_gf, summult, areadif")
      End If

      strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
          pDataset.Tables(0).Rows(iRow)(1) & ", " & _
          pDataset.Tables(0).Rows(iRow)(2) & ", " & _
          pDataset.Tables(0).Rows(iRow)(3) & ", " & _
          pDataset.Tables(0).Rows(iRow)(4) & ")"

      WriteLogInfo(strRow)
      blnFoundOne = True
    Next
    If blnFoundOne = False Then
      WriteLogInfo("0 outliers found")
    End If
    WriteLogInfo("")
    pDataset = Nothing


    "
    " collect and write out top 4 MIN(AREADIF)
    "
    mpProgressWnd.SetStatus(0, 0, "Reporting Summary Stats MAX(SUMMULT)")
    strSQL = "SELECT MTRS, AREA_SUMF, SUMAREA_GF, SUMMULT, AREADIF FROM QA_MTRS_" & strYear & " WHERE
SUMMULT > " & mdMaxMult & " ORDER BY SUMMULT DESC"
    pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
    WriteLogInfo("MAX(SumMult) > " & mdMaxMult)
    blnFoundOne = False
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
      If blnFoundOne = False Then
        WriteLogInfo("MTRS, area_sumf, sumarea_gf, summult, areadif")
      End If
      strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
          pDataset.Tables(0).Rows(iRow)(1) & ", " & _
          pDataset.Tables(0).Rows(iRow)(2) & ", " & _
```

```
                pDataset.Tables(0).Rows(iRow)(3) & ", " & _
                pDataset.Tables(0).Rows(iRow)(4) & ")"

            WriteLogInfo(strRow)
            blnFoundOne = True
        Next
        If blnFoundOne = False Then
            WriteLogInfo("0 outliers found")
        End If
        WriteLogInfo("")
        pDataset = Nothing

        "
        " All CLASS="" should have a zero MULT and AREA_GF
        "
        mpProgressWnd.SetStatus(0, 0, "Inspecting CLASS="")
        strSQL = "SELECT ROW_ID, MTRS, GRIDID, CLASS, AREA_SUMF, AREA_GF, MULT FROM Fractions_MTRS_" & strYear
& " " & _
                "WHERE CLASS = " AND (MULT > 0 OR AREA_GF > 0)"
        pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
        WriteLogInfo("All CLASS="""" should have a zero MULT and AREA_GF")
        blnFoundOne = False
        For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
            If blnFoundOne = False Then
                WriteLogInfo("ROW_ID, MTRS, GRIDID, CLASS, AREA_SUMF, AREA_GF, MULT")
            End If
            strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
                    pDataset.Tables(0).Rows(iRow)(1) & ", " & _
                    pDataset.Tables(0).Rows(iRow)(2) & ", " & _
                    pDataset.Tables(0).Rows(iRow)(3) & ", " & _
                    pDataset.Tables(0).Rows(iRow)(4) & "," & _
                    pDataset.Tables(0).Rows(iRow)(5) & "," & _
                    pDataset.Tables(0).Rows(iRow)(6) & ")"

            WriteLogInfo(strRow)
            blnFoundOne = True
        Next
        If blnFoundOne = False Then
            WriteLogInfo("0 found")
        End If
        WriteLogInfo("")
        pDataset = Nothing

        "
        " All CLASS <> "" should have non zero MULT and AREA_GF
        "
        mpProgressWnd.SetStatus(0, 0, "Inspecting CLASS="")
        strSQL = "SELECT ROW_ID, MTRS, GRIDID, CLASS, AREA_SUMF, AREA_GF, MULT FROM Fractions_MTRS_" & strYear
& " " & _
                "WHERE CLASS <> " AND (MULT = 0 OR AREA_GF = 0)"
        pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Mins")
        WriteLogInfo("All CLASS <> "" should have non zero MULT and AREA_GF")
        blnFoundOne = False
        For iRow = 0 To pDataset.Tables(0).Rows.Count - 1
            If blnFoundOne = False Then
                WriteLogInfo("ROW_ID, MTRS, GRIDID, CLASS, AREA_SUMF, AREA_GF, MULT")
            End If
            strRow = "(" & pDataset.Tables(0).Rows(iRow)(0) & ", " & _
                    pDataset.Tables(0).Rows(iRow)(1) & ", " & _
                    pDataset.Tables(0).Rows(iRow)(2) & ", " & _
                    pDataset.Tables(0).Rows(iRow)(3) & ", " & _
                    pDataset.Tables(0).Rows(iRow)(4) & "," & _
                    pDataset.Tables(0).Rows(iRow)(5) & "," & _
                    pDataset.Tables(0).Rows(iRow)(6) & ")"

            WriteLogInfo(strRow)
            blnFoundOne = True
        Next
        If blnFoundOne = False Then
            WriteLogInfo("0 found")
```

```
    End If
    WriteLogInfo("")
    pDataset = Nothing

    mpProgressWnd.SetStatus(0, 0, "")

End Function

'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
' list the Usage records with no fractions
'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
Public Function ListUsageWithNoFractions(ByVal strYear As String) As Boolean

    WriteLogInfo("Usage records with no fractions: YEAR=" & strYear)
    Dim blnFoundOne As Boolean = False

    ' select the records from the usage database that have polygon
    ' IDs in the specified year column i.e. POLY95
    Dim strSQL As String
    strSQL = "SELECT distinct Poly" & strYear & " " & _
          "FROM MeBr " & _
          "WHERE Poly" & strYear & " > 0"

    Dim pDataset As DataSet
    pDataset = mpDatabase.GetDataset(msUsageDB, strSQL, "UsagePolys")


    ' query the fraction table and find IDs
    Dim iRow As Long
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1

       mpProgressWnd.SetStatus(iRow, pDataset.Tables(0).Rows.Count, "Matching usage records to polygon fractions")

       Dim strICFID As String
       strICFID = pDataset.Tables(0).Rows(iRow)(0)

       Dim strSQL2 As String
       strSQL2 = "SELECT * " & _
             "FROM Fractions_Fields_" & strYear & " " & _
             "WHERE ICF_ID = " & strICFID
       Dim pDataset2 As DataSet
       pDataset2 = mpDatabase.GetDataset(msQADB, strSQL2, "Fractions")

       ' as long as there are records then we are good.  If there are no records
       ' then we have usage records with no fractions - BAD.
       If pDataset2.Tables(0).Rows.Count = 0 Then

          ' well, we need to report this.  I guess we could just write this ICF_ID
          ' to the logfile, but while we have the compiler crankin lets just report
          "the usage records so the QA person has them.

          ' print out a header row for the first one
          If blnFoundOne = False Then
             WriteLogInfo("YEARUSE_NO, UPD_MTRS, APPLIC_DT, GROW_SHORT,SITE_LOC_I, LBS_CHM_US,
ACRE_PLANT, ACRE_TREAT, Poly95, Poly96, Poly97, Poly98, Poly99, poly00, Poly01")
             blnFoundOne = True
          End If

          ' query to get usage records with this ID
          Dim strSql3 As String
          strSql3 = "SELECT YEARUSE_NO, UPD_MTRS, APPLIC_DT, GROW_SHORT,SITE_LOC_I, LBS_CHM_US,
ACRE_PLANT, ACRE_TREAT, Poly95, Poly96, Poly97, Poly98, Poly99, poly00, Poly01  " & _
                "FROM MeBr " & _
                "WHERE Poly" & strYear & " = " & strICFID

          WriteLogInfo("SQL fractions:" & strSQL2)
          WriteLogInfo("SQL usage:" & strSql3)

          Dim pDataset3 As DataSet
          pDataset3 = mpDatabase.GetDataset(msUsageDB, strSql3, "BadRecords")
```

```
            '' loop output and log records
            Dim jRow As Long
            For jRow = 0 To pDataset3.Tables(0).Rows.Count - 1
               Dim strRow As String
               strRow = "(" & pDataset3.Tables(0).Rows(jRow)(0) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(1) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(2) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(3) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(4) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(5) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(6) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(7) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(8) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(9) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(10) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(11) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(12) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(13) & ", " & _
                     pDataset3.Tables(0).Rows(jRow)(14) & ")"

               WriteLogInfo(strRow)

            Next jRow
            pDataset3 = Nothing

         End If

         pDataset2 = Nothing

      Next
      If blnFoundOne = False Then
         WriteLogInfo("0 errors found")
      End If

      WriteLogInfo("")

      pDataset = Nothing

      mpProgressWnd.SetStatus(0, 0, "")

   End Function

   ''*************************************************************************
   '' open a text file, read in and count the rows
   ''*************************************************************************
   Public Function WriteLogInfo(ByVal strLogMessage As String)


      Dim pStreamWriter As System.IO.StreamWriter
      pStreamWriter = New System.IO.StreamWriter(msQALogfile, True)
      If pStreamWriter Is Nothing Then
         MsgBox("ERROR writeing to log file: "' & msQALogfile & """)
         Exit Function
      End If

      pStreamWriter.WriteLine(strLogMessage)
      pStreamWriter.Flush()
      pStreamWriter.Close()
      pStreamWriter = Nothing

   End Function


   ''*************************************************************************
   '' list the Usage records with no fractions
   ''*************************************************************************
   Public Function ListPolygonsWithNoUsage(ByVal strYear As String) As Boolean

      WriteLogInfo("Polygons with no usage records: YEAR=" & strYear)
```

```
    Dim blnFoundOne As Boolean = False

    " select the records from the usage database that have polygon
    " IDs in the specified year column i.e. POLY95
    Dim strSQL As String
    strSQL = "SELECT distinct ICF_ID " & _
            "FROM Fractions_Fields_" & strYear

    Dim pDataset As DataSet
    pDataset = mpDatabase.GetDataset(msQADB, strSQL, "Polygons")

    '"obtain connection that we can use across updates
    Dim pConnection As OleDbConnection = mpDatabase.GetConnection(msUsageDB)

    " loop polygons
    Dim iRow As Long
    For iRow = 0 To pDataset.Tables(0).Rows.Count - 1

        mpProgressWnd.SetStatus(iRow, pDataset.Tables(0).Rows.Count, "Matching polygon fractions to usage records")

        Dim strICFID As String
        strICFID = pDataset.Tables(0).Rows(iRow)(0)

        Dim strSQL2 As String
        strSQL2 = "SELECT yearuse_no " & _
                "FROM MeBr " & _
                "WHERE Poly" & strYear & " = " & strICFID
        Dim pDataset2 As DataSet
        pDataset2 = mpDatabase.GetDataset(pConnection, strSQL2, "Usage")

        " as long as there are records then we are good.  If there are no records
        " then we have a polygon without usage.  This isn't always bad, but we need to list them
        If pDataset2.Tables(0).Rows.Count = 0 Then


            " print out a header row for the first one
            If blnFoundOne = False Then
                WriteLogInfo("ICF_ID")
                blnFoundOne = True
            End If

            " TODO()
            " insert code to get polygon data from map layer
            " fraction file data doesn't tell us much
            WriteLogInfo(strICFID)


        End If
        pDataset2 = Nothing


    Next
    pConnection.Close()
    pConnection = Nothing

    If blnFoundOne = False Then
        WriteLogInfo("0 errors found")
    End If
    WriteLogInfo("")

    pDataset = Nothing

    mpProgressWnd.SetStatus(0, 0, "")
End Function

'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
" Check MTRS Usage against MTRS Fractions
'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
Public Function CheckMTRSUsageagainstMTRSFractions(ByVal strYear As String) As Boolean
```

```
        WriteLogInfo("List MTRS with usage and no fractions: YEAR=" & strYear)
        Dim blnFoundOne As Boolean = False

        " MTRS sections from database
        'Kern, Benito, Santa Clara, Monterey, Santa Cruz
        Dim strMTRSWhereClause As String
        'strMTRSWhereClause = " WHERE COUNTY_CD in ('44','27','43','35','15')"
        strMTRSWhereClause = pAppConfig.GetConfigItem("AppConfiguration/MTRSWhereClause")

        Dim strSQL As String
        strSQL = "SELECT distinct UPD_MTRS from MeBr " & strMTRSWhereClause

        Dim pDataset As DataSet
        pDataset = mpDatabase.GetDataset(msUsageDB, strSQL, "UsageMTRS")

        '"obtain connection that we can use across updates
        Dim pConnection As OleDbConnection = mpDatabase.GetConnection(msQADB)

        " loop MTRS
        Dim iRow As Long
        For iRow = 0 To pDataset.Tables(0).Rows.Count - 1

            mpProgressWnd.SetStatus(iRow, pDataset.Tables(0).Rows.Count, "Check MTRS Usage against MTRS Fractions")
            Dim strMTRS As String
            strMTRS = "" & pDataset.Tables(0).Rows(iRow)(0)

            Dim strSQL2 As String
            strSQL2 = "SELECT distinct MTRS " & _
                    "FROM Fractions_MTRS_" & strYear & " " & _
                    "WHERE MTRS = '" & strMTRS & "'"

            Dim pDataset2 As DataSet
            pDataset2 = mpDatabase.GetDataset(pConnection, strSQL2, "Usage")

            " as long as there are records then we are good.  If there are no records
            " then we have a polygon without usage.  This isn't always bad, but we need to list them
            If pDataset2.Tables(0).Rows.Count = 0 Then

                blnFoundOne = True

                If strMTRS = "" Then
                    strMTRS = "null"
                End If
                " report the MTRS with usage that has no fraction
                WriteLogInfo(strMTRS)

            End If
            pDataset2 = Nothing
        Next
        pConnection.Close()
        pConnection = Nothing

        If blnFoundOne = False Then
            WriteLogInfo("0 errors found")
        End If
        WriteLogInfo("")
        pDataset = Nothing
        mpProgressWnd.SetStatus(0, 0, "")

    End Function

End Class
```

StatusWnd.vb

```
Public Class progressWnd
    Inherits System.Windows.Forms.Form

#Region " Windows Form Designer generated code "
```

```
Public Sub New()
    MyBase.New()

    'This call is required by the Windows Form Designer.
    InitializeComponent()

    'Add any initialization after the InitializeComponent() call

End Sub

'Form overrides dispose to clean up the component list.
Protected Overloads Overrides Sub Dispose(ByVal disposing As Boolean)
    If disposing Then
        If Not (components Is Nothing) Then
            components.Dispose()
        End If
    End If
    MyBase.Dispose(disposing)
End Sub

'Required by the Windows Form Designer
Private components As System.ComponentModel.IContainer

'NOTE: The following procedure is required by the Windows Form Designer
'It can be modified using the Windows Form Designer.
'Do not modify it using the code editor.
Friend WithEvents ProgressBar1 As System.Windows.Forms.ProgressBar
Friend WithEvents Label1 As System.Windows.Forms.Label
Friend WithEvents Label2 As System.Windows.Forms.Label
<System.Diagnostics.DebuggerStepThrough()> Private Sub InitializeComponent()
    Me.ProgressBar1 = New System.Windows.Forms.ProgressBar()
    Me.Label1 = New System.Windows.Forms.Label()
    Me.Label2 = New System.Windows.Forms.Label()
    Me.SuspendLayout()
    '
    'ProgressBar1
    '
    Me.ProgressBar1.Anchor = ((System.Windows.Forms.AnchorStyles.Top Or System.Windows.Forms.AnchorStyles.Left) _
                Or System.Windows.Forms.AnchorStyles.Right)
    Me.ProgressBar1.Location = New System.Drawing.Point(16, 24)
    Me.ProgressBar1.Name = "ProgressBar1"
    Me.ProgressBar1.Size = New System.Drawing.Size(280, 23)
    Me.ProgressBar1.Step = 1
    Me.ProgressBar1.TabIndex = 0
    '
    'Label1
    '
    Me.Label1.Location = New System.Drawing.Point(16, 72)
    Me.Label1.Name = "Label1"
    Me.Label1.Size = New System.Drawing.Size(280, 24)
    Me.Label1.TabIndex = 1
    Me.Label1.Text = "Label1"
    Me.Label1.TextAlign = System.Drawing.ContentAlignment.MiddleCenter
    '
    'Label2
    '
    Me.Label2.Location = New System.Drawing.Point(16, 48)
    Me.Label2.Name = "Label2"
    Me.Label2.Size = New System.Drawing.Size(280, 16)
    Me.Label2.TabIndex = 2
    Me.Label2.Text = "Label2"
    Me.Label2.TextAlign = System.Drawing.ContentAlignment.MiddleCenter
    '
    'progressWnd
    '
    Me.AutoScaleBaseSize = New System.Drawing.Size(5, 13)
    Me.ClientSize = New System.Drawing.Size(312, 101)
    Me.Controls.AddRange(New System.Windows.Forms.Control() {Me.Label2, Me.Label1, Me.ProgressBar1})
    Me.Name = "progressWnd"
```

```
        Me.Text = "Progress"
        Me.ResumeLayout(False)

    End Sub

#End Region

    Public Function SetStatus(ByVal lngCurrent As Long, ByVal lngTotal As Long, ByVal strMessage As String)
        Label1.Text = strMessage
        If lngTotal <> 0 Then
            ProgressBar1.Value = lngCurrent / lngTotal * 100
        End If
        Label2.Text = lngCurrent & " of " & lngTotal

        Label2.Refresh()
        Label1.Refresh()
        ProgressBar1.Refresh()
        Me.Refresh()

    End Function
End Class
```

TestFieldFractions.vb

```
''*********************************************************************************************
'' Angelita C. - California Schools Project
'' Created by: Matt Werth, ICF Consulting GIS Dept
'' 1/8/04
''
'' This class will test the FIELDS fraction files procuced for the Angelita C project.
''
''*********************************************************************************************
Imports System.Runtime.InteropServices
Imports ESRI.ArcObjects.Core
Imports ESRI.ArcObjects.Samples.CatIDs
Imports ESRI.ArcObjects.Samples.BaseClasses

<ComClass(TestFieldFractions.ClassId, TestFieldFractions.InterfaceId, TestFieldFractions.EventsId)> _
Public NotInheritable Class TestFieldFractions
    Inherits BaseCommand

    Private mpMxApp As IMxApplication

#Region "COM GUIDs"
    ' ** THESE WERE CREATED ALONG WHEN WE "ADD COM CLASS"
    'These  GUIDs provide the COM identity for this class
    ' and its COM interfaces. If you change them, existing
    ' clients will no longer be able to access the class.
    Public Const ClassId As String = "1D930377-EF80-4048-9F3D-4F414374451C"
    Public Const InterfaceId As String = "4517C70A-CC48-46DA-9B48-F76E216C512C"
    Public Const EventsId As String = "72FDE6FC-6175-4AF2-9043-A27D6B4DF64C"
#End Region

#Region "Register Unregister Component"
    <ComRegisterFunction(), ComVisible(False)> _
      Public Shared Sub RegisterFunction(ByVal t As Type)
        If Not (t Is Nothing) Then
            MxCommand.Register("HKEY_CLASSES_ROOT\CLSID\{" + t.GUID.ToString() + "}")
        End If
    End Sub

    <ComUnregisterFunction(), ComVisible(False)> _
     Public Shared Sub UnregisterFunction(ByVal t As Type)
        If Not (t Is Nothing) Then
            MxCommand.Unregister("HKEY_CLASSES_ROOT\CLSID\{" + t.GUID.ToString() + "}")
        End If
    End Sub

    '<ComRegisterFunction(), ComVisible(False)> _
```

```
    'Public Shared Sub OnRegister(ByVal regKey As String)
    '    MxCommand.Register(regKey)
    'End Sub

    '<ComUnregisterFunction(), ComVisible(False)> _
    'Public Shared Sub OnUnRegister(ByVal regKey As String)
    '    MxCommand.Unregister(regKey)
    'End Sub
#End Region

    ' A creatable COM class must have a Public Sub New()
    ' with no parameters, otherwise, the class will not be
    ' registered in the COM registry and cannot be created
    ' via CreateObject.
    Public Sub New()
        MyBase.New()

        MyBase.m_category = "ICF CA_Schools"
        MyBase.m_caption = "Test Fields Fractions"
        MyBase.m_message = "Test Fields Fractions"
        MyBase.m_toolTip = "Test Fields Fractions (VB.NET)"
        MyBase.m_name = "ICF CA Schools_FrationTester VB.NET"

        MyBase.m_bitmap = New System.Drawing.Bitmap _
          ((GetType(TestFieldFractions).Assembly.GetManifestResourceStream _
          ("FractionTester.database_connection_with_check.bmp")))

    End Sub

    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    Public Overrides Sub OnCreate(ByVal hook As Object)
        If Not (hook Is Nothing) Then
            If TypeOf (hook) Is IMxApplication Then
                mpMxApp = CType(hook, IApplication)
            End If
        End If
    End Sub

    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    Public Overrides Sub OnClick()
        TestFieldFractions()
    End Sub
    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    '
    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    Public Function LoadConfig()
    End Function
    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    '' main entry function called from the command OnClick()
    '''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    Public Function TestFieldFractions()

        ''Dim mxDoc As IMxDocument = CType(mpMxApp.Document, IMxDocument)
        ''Dim activeView As IActiveView = mxDoc.ActiveView
        ''activeView.Extent = activeView.FullExtent
        ''activeView.Refresh()

        Dim strYear As String = "00"

        '' push fraction file into database
        'Upload_FractionFile_To_Database(strYear)

        '' create & populate stats table and populate with groupable sums

    End Function

End Class
```

APPENDIX O. SAMPLE QUALITY ASSURANCE TABLE

```
**************************************************************************
  Angelita C project, for EPA, by ICF consulting
  Fraction File QA: 1/27/2004 7:31:28 PM
**************************************************************************
**************************************
  MTRS Fractions for year 00
  1/27/2004 7:31:28 PM
**************************************
MIN(AreaDif) < -0.0001
0 outliers found

MAX(AreaDif) > 0.0001
0 outliers found

MIN(SumMult) < 0.99999
0 outliers found

MAX(SumMult) > 1.00001
0 outliers found

All CLASS="" should have a zero MULT and AREA_GF
0 found

All CLASS <> " should have non zero MULT and AREA_GF
0 found


List MTRS with usage and no fractions: YEAR=00
null
......
00
000000000
M??14?08
M11?11E02

**************************************
  FIELD Fractions for year 00
  1/27/2004 7:32:43 PM
**************************************
MIN(AreaDif) < -0.0001
0 outliers found

MAX(AreaDif) > 0.0001
0 outliers found

MIN(SumMult) < 0.99999
0 outliers found

MAX(SumMult) > 1.00001
0 outliers found

Usage records with no fractions: YEAR=00
```

```
0 errors found

Polygons with no usage records: YEAR=00
ICF_ID
23
27
28
20001
20002
20003
20004
20006
20007
20008
20010
20011
20012
20013
20014
20016
20017
20019
20020
20021
20022
20023


****************************************
  MTRS Fractions for year 01
  1/27/2004 7:33:29 PM
****************************************
MIN(AreaDif) < -0.0001
0 outliers found

MAX(AreaDif) > 0.0001
0 outliers found

MIN(SumMult) < 0.99999
0 outliers found

MAX(SumMult) > 1.00001
0 outliers found

All CLASS="" should have a zero MULT and AREA_GF
0 found

All CLASS <> " should have non zero MULT and AREA_GF
0 found


List MTRS with usage and no fractions: YEAR=01
null
......
00
000000000
M??14?08
```

```
M11?11E02

**************************************
  FIELD Fractions for year 01
  1/27/2004 7:34:44 PM
**************************************
MIN(AreaDif) < -0.0001
0 outliers found

MAX(AreaDif) > 0.0001
0 outliers found

MIN(SumMult) < 0.99999
0 outliers found

MAX(SumMult) > 1.00001
0 outliers found

Usage records with no fractions: YEAR=01
YEARUSE_NO, UPD_MTRS, APPLIC_DT, GROW_SHORT,SITE_LOC_I, LBS_CHM_US,
ACRE_PLANT, ACRE_TREAT, Poly95, Poly96, Poly97, Poly98, Poly99, poly00,
Poly01
(99939531, M12S02E22, 9/16/1999, 440447A, 010001, 3060.8999, 17, 15.5, , , ,
, 1, 1)

Polygons with no usage records: YEAR=01
ICF_ID
1077
20001
20002
20003
20004
20006
20007
20008
20010
20011
20012
20013
20014
20015
20016
20017
20019
20020
20021
20022
20023
```

APPENDIX P. ARCOBJECTS SCRIPT FOR DEVELOPING DISTANCE AND ANGLE TABLE

DistanceAngle.cls

```
VERSION 1.0 CLASS
BEGIN
  MultiUse = -1  'True
END
Attribute VB_Name = "ThisDocument"
Attribute VB_GlobalNameSpace = False
Attribute VB_Creatable = False
Attribute VB_PredeclaredId = True
Attribute VB_Exposed = True
Option Explicit

Private Sub UIButtonControl1_Click()

        '    PointDistance "P:\CA_Schools\StateRun\s_alb.shp", _
        '                "c:\projects\tmp\g_pnt", _
        '                42240, _
        '                "P:\CA_Schools\StateRun\DistAngle.shp"
          DistanceAngleToText "P:\CA_Schools\StateRun\schools_11132003_alb.shp", _
                "P:\CA_Schools\StateRun\g_pnt", _
                42240, _
                "P:\CA_Schools\StateRun\DistAngle"

End Sub

''*********************************************************************************************
'' PointDistance - took 8 hours
'' determine the distance between each point in the input
'' SHP to every point within X in target SHP
''*********************************************************************************************
Public Sub DistanceAngle(strFromSHP As String, strToShp As String, dlbRadius As Double, strOutputShp As String)

          Dim intTotalSchools As Integer
          Dim intCurSchool As Integer
          Dim pProgressWindow As New frmProgress
          ''
          '' open the From  layer
          ''
          Dim pFromFeatureClass As IFeatureClass
          Set pFromFeatureClass = OpenShapefile(strFromSHP)

          intTotalSchools = pFromFeatureClass.FeatureCount(Nothing)

          Dim iCol_SiteId As Integer
          iCol_SiteId = pFromFeatureClass.FindField("SCHOOL_ID")

          '' open the To layer
          Dim pToFeatureClass As IFeatureClass
          Set pToFeatureClass = OpenCoverage(strToShp)
          ''Set pToFeatureClass = OpenShapefile(strToShp)

          Dim iCol_GridId As Integer
          iCol_GridId = pToFeatureClass.FindField("GRIDID")
          ''
          '' create output shapefile
          ''
          '' change input featureclass into geodataset so we can access the input spatial reference
          Dim pGeoDataset As IGeoDataset
          Set pGeoDataset = pFromFeatureClass

          Dim pOutputFClass As IFeatureClass
          Set pOutputFClass = CreateOutputShapfile(strOutputShp, pGeoDataset.SpatialReference)
```

```
''
'' Do some setup stuff that should be done once outside the main loop
''
    '' fire up an output cursor to write the output line to
    Dim pOutInsertCursor As IFeatureCursor
    Set pOutInsertCursor = pOutputFClass.Insert(True)

    ''
    '' setup the spatial reference that we can use to reproject the line
    '' so the we can get a good angle calculation between the school and grid cell
    ''
    Dim pSRF As ISpatialReferenceFactory2
    Set pSRF = New SpatialReferenceEnvironment

    Dim pGCS As IGeographicCoordinateSystem
    Set pGCS = pSRF.CreateGeographicCoordinateSystem(esriSRGeoCS_NAD1983)

    Dim pGeoNad83SpatialRef As ISpatialReference
    Set pGeoNad83SpatialRef = pGCS
''
'' loop From Layer
''   select ToLayer object within buffer dist of single FromLayer obj
''   output line object for each ToLayer obj
''

'' show the progresswindow
pProgressWindow.Show vbModeless

'' get a cursor of all objs in FromShp
Dim pFeatureCursor As IFeatureCursor
Set pFeatureCursor = pFromFeatureClass.Search(Nothing, True)

'' initialize the cursor
Dim pFromFeature As IFeature
Set pFromFeature = pFeatureCursor.NextFeature

'' loop the cursor (entire FromShp)
Do Until pFromFeature Is Nothing

    '' check to see if user requested a cancel
    DoEvents
    If pProgressWindow.blnCancelPressed = True Then
        Unload pProgressWindow
        Set pProgressWindow = Nothing
        MsgBox "Quitting"
        Exit Do ' bail
    End If

    '' get a topo operator for the current feature obj
    Dim pTopoOp As ITopologicalOperator
    Set pTopoOp = pFromFeature.Shape

    '' update the prgress counter
    intCurSchool = intCurSchool + 1
    pProgressWindow.UpdateProgress intCurSchool, intTotalSchools

    '' create a spatial filter that is a buffer aroung the current FromFeature
    Dim pSpatialFilt As ISpatialFilter
    Set pSpatialFilt = New SpatialFilter
    Set pSpatialFilt.Geometry = pTopoOp.buffer(dlbRadius)
    'pSpatialFilt.SpatialRel = esriSpatialRelWithin
    pSpatialFilt.SpatialRel = esriSpatialRelContains

    Dim pToCursor As IFeatureCursor
    Set pToCursor = pToFeatureClass.Search(pSpatialFilt, True)

    '' loop the ToShp features within buffer of current FromShp feature
```

```
    Dim pToFeature As IFeature
    Set pToFeature = pToCursor.NextFeature
    Do Until pToFeature Is Nothing


      Dim pNewFeature As IFeature
      Dim pNewFeatureBuffer As IFeatureBuffer
      Set pNewFeatureBuffer = pOutputFClass.CreateFeatureBuffer
      Set pNewFeature = pNewFeatureBuffer

      '' create the shape geometry
      Dim pPolyline As IPolyline
      Set pPolyline = New Polyline
      pPolyline.FromPoint = pFromFeature.Shape
      pPolyline.ToPoint = pToFeature.Shape
      Set pPolyline.SpatialReference = pFromFeature.Shape.SpatialReference

      '' set the new lines shape to the line we just created
      Set pNewFeature.Shape = pPolyline

      ''
      '' set the data field values
      ''
      pNewFeature.Value(2) = pFromFeature.Value(iCol_SiteId) '' SITEID / SCHOOL_ID
      pNewFeature.Value(3) = pToFeature.Value(iCol_GridId) '' GRIDID
'     pNewFeature.Value(4) = pPolyline.FromPoint.x '' Sx
'     pNewFeature.Value(5) = pPolyline.FromPoint.Y '' Sy
'     pNewFeature.Value(6) = pPolyline.ToPoint.x '' Gx
'     pNewFeature.Value(7) = pPolyline.ToPoint.Y '' Gy
      pNewFeature.Value(4) = pPolyline.Length '' Distance

      '' reproject the line to Geographic so we can get an angle
      pPolyline.Project pGeoNad83SpatialRef
      Dim pLine As ILine
      Set pLine = New Line
      pLine.FromPoint = pPolyline.FromPoint
      pLine.ToPoint = pPolyline.ToPoint

      Dim pi As Double
      pi = 4 * Atn(1)

      Dim dblAngleDeg As Double
      dblAngleDeg = pLine.Angle * 360 / (2 * pi)

      '' change angle to be zero north clockwise
      dblAngleDeg = 450 - dblAngleDeg
      If dblAngleDeg > 360 Then
        dblAngleDeg = dblAngleDeg - 360
      End If

      '' finally store the angle in the new feature
      pNewFeature.Value(5) = dblAngleDeg '' Angle


      '' now add the a line object to output shp file
      Dim lngID As Long
      lngID = pOutInsertCursor.InsertFeature(pNewFeatureBuffer)

      Set pToFeature = pToCursor.NextFeature

    Loop

    Set pToCursor = Nothing
    Set pFromFeature = pFeatureCursor.NextFeature
  Loop

  'Flush the feature cursor to the database
  pOutInsertCursor.Flush
  Unload pProgressWindow
  Set pProgressWindow = Nothing
```

```
End Sub

''*************************************************************************************************
'' PointDistance
'' determine the distance between each point in the input
'' SHP to every point within X in target SHP
''*************************************************************************************************
Public Sub DistanceAngleToText(strFromSHP As String, strToShp As String, dlbRadius As Double, strOutputFile As String)


    Dim intTotalSchools As Integer
    Dim intCurSchool As Integer
    Dim pProgressWindow As New frmProgress


    ''
    '' open the From  layer
    ''
    Dim pFromFeatureClass As IFeatureClass
    Set pFromFeatureClass = OpenShapefile(strFromSHP)

    intTotalSchools = pFromFeatureClass.FeatureCount(Nothing)

    Dim iCol_SiteId As Integer
    iCol_SiteId = pFromFeatureClass.FindField("SCHOOL_ID")

    '' open the To layer
    Dim pToFeatureClass As IFeatureClass
    Set pToFeatureClass = OpenCoverage(strToShp)
    ''Set pToFeatureClass = OpenShapefile(strToShp)

    Dim iCol_GridId As Integer
    iCol_GridId = pToFeatureClass.FindField("GRIDID")

    ''
    '' create output textfile
    ''
    ''
    Dim intFileCount As Integer
    intFileCount = 1

    Dim pOutFile As Scripting.TextStream
    Dim pFSO As Scripting.FileSystemObject
    Set pFSO = New FileSystemObject
    On Error Resume Next
    pFSO.DeleteFile strOutputFile & "_" & intFileCount & ".txt"
    On Error GoTo 0

    ''
    '' create output shapefile
    ''
    '' change input featureclass into geodataset so we can access the input spatial reference
    Dim pGeoDataset As IGeoDataset
    Set pGeoDataset = pFromFeatureClass

    Dim pOutputFClass As IFeatureClass
    Set pOutputFClass = CreateOutputShapfile("C:\temp\Temp.shp", pGeoDataset.SpatialReference)

    ''
    '' Do some setup stuff that should be done once outside the main loop
    ''
       Dim lngRowId As Long



       ''
       '' setup the spatial reference that we can use to reproject the line
       '' so the we can get a good angle calculation between the school and grid cell
       ''
       Dim pSRF As ISpatialReferenceFactory2
```

```
    Set pSRF = New SpatialReferenceEnvironment

    Dim pGCS As IGeographicCoordinateSystem
    Set pGCS = pSRF.CreateGeographicCoordinateSystem(esriSRGeoCS_NAD1983)

    Dim pGeoNad83SpatialRef As ISpatialReference
    Set pGeoNad83SpatialRef = pGCS

"
" loop From Layer
"   select ToLayer object within buffer dist of single FromLayer obj
"   output line object for each ToLayer obj
"

" show the progresswindow
pProgressWindow.Show vbModeless

" get a cursor of all objs in FromShp
Dim pFeatureCursor As IFeatureCursor
Set pFeatureCursor = pFromFeatureClass.Search(Nothing, True)

" initialize the cursor
Dim pFromFeature As IFeature
Set pFromFeature = pFeatureCursor.NextFeature

" loop the cursor (entire FromShp)
Do Until pFromFeature Is Nothing

    intCurSchool = intCurSchool + 1
    If intCurSchool Mod 1000 = 0 Then
        " every 1000 schools we bump up the output file number
        intFileCount = intFileCount + 1
    End If

    " check to see if user requested a cancel
    DoEvents
    If pProgressWindow.blnCancelPressed = True Then
        Unload pProgressWindow
        Set pProgressWindow = Nothing
        MsgBox "Quitting"
        Exit Do ' bail
    End If

    " get a topo operator for the current feature obj
    Dim pTopoOp As ITopologicalOperator
    Set pTopoOp = pFromFeature.Shape

    " update the prgress counter
    pProgressWindow.UpdateProgress intCurSchool, intTotalSchools

    " create a spatial filter that is a buffer aroung the current FromFeature
    Dim pSpatialFilt As ISpatialFilter
    Set pSpatialFilt = New SpatialFilter
    Set pSpatialFilt.Geometry = pTopoOp.buffer(dlbRadius)
    'pSpatialFilt.SpatialRel = esriSpatialRelWithin
    pSpatialFilt.SpatialRel = esriSpatialRelContains

    Dim pToCursor As IFeatureCursor
    Set pToCursor = pToFeatureClass.Search(pSpatialFilt, True)

    Dim strSiteId As String
    Dim strGridId As String
    Dim strDist As String
    Dim strAngle As String

    " open the file and save it for every school
    Set pOutFile = pFSO.OpenTextFile(strOutputFile & "_" & intFileCount & ".txt", ForAppending, True)


    " loop the ToShp features within buffer of current FromShp feature
```

```
    Dim pToFeature As IFeature
    Set pToFeature = pToCursor.NextFeature
    Do Until pToFeature Is Nothing


      Dim pNewFeature As IFeature
      Dim pNewFeatureBuffer As IFeatureBuffer
      Set pNewFeatureBuffer = pOutputFClass.CreateFeatureBuffer
      Set pNewFeature = pNewFeatureBuffer

      '' create the shape geometry
      Dim pPolyline As IPolyline
      Set pPolyline = New Polyline
      pPolyline.FromPoint = pFromFeature.Shape
      pPolyline.ToPoint = pToFeature.Shape
      Set pPolyline.SpatialReference = pFromFeature.Shape.SpatialReference

      '' set the new lines shape to the line we just created
      Set pNewFeature.Shape = pPolyline

      ''
      '' set the data field values
      ''
      strSiteId = pFromFeature.Value(iCol_SiteId) '' SITEID / SCHOOL_ID
      strGridId = pToFeature.Value(iCol_GridId) '' GRIDID
'      pNewFeature.Value(4) = pPolyline.FromPoint.x '' Sx
'      pNewFeature.Value(5) = pPolyline.FromPoint.Y '' Sy
'      pNewFeature.Value(6) = pPolyline.ToPoint.x '' Gx
'      pNewFeature.Value(7) = pPolyline.ToPoint.Y '' Gy
      strDist = Format(pPolyline.Length, "#######.##") '' Distance

      '' reproject the line to Geographic so we can get an angle
      pPolyline.Project pGeoNad83SpatialRef
      Dim pLine As ILine
      Set pLine = New Line
      pLine.FromPoint = pPolyline.FromPoint
      pLine.ToPoint = pPolyline.ToPoint

      Dim pi As Double
      pi = 4 * Atn(1)

      Dim dblAngleDeg As Double
      dblAngleDeg = pLine.Angle * 360 / (2 * pi)

      '' change angle to be zero north clockwise
      dblAngleDeg = 450 - dblAngleDeg
      If dblAngleDeg > 360 Then
        dblAngleDeg = dblAngleDeg - 360
      End If

      '' finally store the angle in the new feature
      strAngle = Format(dblAngleDeg, "####.##") '' Angle


      '' now add the a line object to output shp file
      lngRowId = lngRowId + 1
      pOutFile.WriteLine lngRowId & "," & strSiteId & "," & strGridId & "," & strDist & "," & strAngle

      Set pToFeature = pToCursor.NextFeature


    Loop

    pOutFile.Close

    Set pToCursor = Nothing
    Set pFromFeature = pFeatureCursor.NextFeature

  Loop
```

```
        'Flush the feature cursor to the database
        ''pOutInsertCursor.Flush

        Unload pProgressWindow
    Set pProgressWindow = Nothing

End Sub

''*************************************************************************************************
'' create output shapefile
''*************************************************************************************************
'Public Function CreateOutputShapfile(strShapefile As String, pInFeatureClass As IFeatureClass) As IFeatureClass
Public Function CreateOutputShapfile(strShapefile As String, pSpatialRef As ISpatialReference) As IFeatureClass

        Const strShapeField As String = "Shape"

        '' delete the file if it already exists
        DeleteShapfile strShapefile

        Dim strPath As String
        Dim strFile As String
        Dim strExtension As String
        SplitPath strShapefile, strPath, strFile, strExtension


        Dim pShapefileWSF As IWorkspaceFactory
        Set pShapefileWSF = New ShapefileWorkspaceFactory

        Dim pFWorkspace As IFeatureWorkspace
        Set pFWorkspace = pShapefileWSF.OpenFromFile(strPath, 0)


        ' Set up a simple fields collection
        Dim pFields As IFields
        Dim pFieldsEdit As IFieldsEdit
        Set pFields = New esriCore.Fields
        Set pFieldsEdit = pFields

        Dim pField As Field
        Dim pFieldEdit As IFieldEdit

        '' (RowId, [SHAPE], SiteId, GridId, Sx, Sy, Gx, Gy, Dist, Ang)

        ''
        '' SHAPE field
        ''
        Set pField = New esriCore.Field
        Set pFieldEdit = pField
        pFieldEdit.Name = strShapeField
        pFieldEdit.Type = esriFieldTypeGeometry

        Dim pGeomDef As IGeometryDef
        Dim pGeomDefEdit As IGeometryDefEdit
        Set pGeomDef = New GeometryDef
        Set pGeomDefEdit = pGeomDef

        pGeomDefEdit.GeometryType = esriGeometryPolyline
        Set pGeomDefEdit.SpatialReference = pSpatialRef
        'Set .SpatialReference = New UnknownCoordinateSystem

        Set pFieldEdit.GeometryDef = pGeomDef
        pFieldsEdit.AddField pField

        Set pField = New esriCore.Field
        Set pFieldEdit = pField
        pFieldEdit.Name = "SiteId"
        pFieldEdit.Type = esriFieldTypeInteger
        pFieldsEdit.AddField pField

        Set pField = New esriCore.Field
```

```
        Set pFieldEdit = pField
        pFieldEdit.Name = "GridId"
        pFieldEdit.Type = esriFieldTypeInteger
        pFieldsEdit.AddField pField

    '   Set pField = New esriCore.Field
    '   Set pFieldEdit = pField
    '   pFieldEdit.Name = "Sx"
    '   pFieldEdit.Type = esriFieldTypeDouble
    '   pFieldsEdit.AddField pField
    '
    '   Set pField = New esriCore.Field
    '   Set pFieldEdit = pField
    '   pFieldEdit.Name = "Sy"
    '   pFieldEdit.Type = esriFieldTypeDouble
    '   pFieldsEdit.AddField pField
    '
    '   Set pField = New esriCore.Field
    '   Set pFieldEdit = pField
    '   pFieldEdit.Name = "Gx"
    '   pFieldEdit.Type = esriFieldTypeDouble
    '   pFieldsEdit.AddField pField
    '
    '   Set pField = New esriCore.Field
    '   Set pFieldEdit = pField
    '   pFieldEdit.Name = "Gy"
    '   pFieldEdit.Type = esriFieldTypeDouble
    '   pFieldsEdit.AddField pField

        Set pField = New esriCore.Field
        Set pFieldEdit = pField
        pFieldEdit.Name = "Dist"
        pFieldEdit.Type = esriFieldTypeDouble
        pFieldsEdit.AddField pField

        Set pField = New esriCore.Field
        Set pFieldEdit = pField
        pFieldEdit.Name = "Ang"
        pFieldEdit.Type = esriFieldTypeDouble
        pFieldsEdit.AddField pField

    On Error GoTo ShpError

        Dim pFClass As IFeatureClass
        Set pFClass = pFWorkspace.CreateFeatureClass(strFile & strExtension, pFields, Nothing, Nothing, esriFTSimpleEdge,
    strShapeField, "")

        Dim x As Integer
        x = 12 * 34

        'Dim pGeoDataset As IGeoDataset
        'Set pGeoDataset = pFClass
        'Set pGeoDataset.SpatialReference = pSpatialRef

        ''pFClass.
        '' return the new shapefile
        Set CreateOutputShapfile = pFClass
        Exit Function

    ShpError:
    MsgBox "Error: " & Err.Description
        Set CreateOutputShapfile = pFClass
End Function



''*************************************************************************************************
'' split a full path into path and filename
''*************************************************************************************************
```

```
Public Function SplitPath(ByVal strInPath As String, ByRef strPath As String, ByRef strFile As String, ByRef strExtension As String)

  Dim iSlash As Integer
  iSlash = InStrRev(strInPath, "\")
  strPath = Left(strInPath, iSlash - 1)
  strFile = Mid(strInPath, iSlash + 1, Len(strInPath) - iSlash - 4)
  strExtension = Right(strInPath, 4)

End Function

''****************************************************************************************************
'' Delete shapefile (if it exists)
''****************************************************************************************************
Public Function DeleteShapfile(strShapefile As String)


          Dim sPath As String
          Dim sFile As String
          Dim sExt As String
          SplitPath strShapefile, sPath, sFile, sExt

          Dim sarFilenames() As String

          Dim pFUtil As New clsFileUtil
          Dim lngCount As Long
          lngCount = pFUtil.FindAllFilesInPath(sPath, sFile & ".*", sarFilenames, False, False)

          If lngCount = 0 Then
            Exit Function
          End If

          Dim sCurFile As String

          Dim pFSO As New FileSystemObject
          Dim i As Integer
          For i = 0 To UBound(sarFilenames)
            sCurFile = sarFilenames(i)
            pFSO.DeleteFile sPath & sCurFile, True
          Next

          Set pFSO = Nothing
          Set pFUtil = Nothing

End Function


''****************************************************************************************************
'' open a shapefile and return the feature class
''****************************************************************************************************
Public Function OpenShapefile(strShapefile As String) As FeatureClass

          '' assume the path includes the shapefile name
          Dim strPath As String
          Dim strFile As String

          Dim iSlash As Integer
          iSlash = InStrRev(strShapefile, "\")
          strPath = Left(strShapefile, iSlash - 1)
          strFile = Right(strShapefile, Len(strShapefile) - iSlash)


          Dim pWSFactory As IWorkspaceFactory
          Dim pWorkspace As IFeatureWorkspace

          Set pWSFactory = New ShapefileWorkspaceFactory
          Set pWorkspace = pWSFactory.OpenFromFile(strPath, 0)

          Dim pFClass As IFeatureClass
          Set pFClass = pWorkspace.OpenFeatureClass(strFile)
```

```
        Set OpenShapefile = pFClass

End Function


''*********************************************************************************************
'' open a Coverage and return the feature class
''*********************************************************************************************
Public Function OpenCoverage(strCoverage As String) As FeatureClass

        '' assume the path includes the shapefile name
        Dim strPath As String
        Dim strFile As String

        Dim iSlash As Integer
        iSlash = InStrRev(strCoverage, "\")
        strPath = Left(strCoverage, iSlash - 1)
        strFile = Right(strCoverage, Len(strCoverage) - iSlash)


        Dim pWSFactory As IWorkspaceFactory
        Dim pWorkspace As IFeatureWorkspace

        Set pWSFactory = New ArcInfoWorkspaceFactory
        Set pWorkspace = pWSFactory.OpenFromFile(strPath, 0)

        Dim pFClass As IFeatureClass
        Set pFClass = pWorkspace.OpenFeatureClass(strFile & ":point")

        Set OpenCoverage = pFClass

End Function
```

ClsFileUtil.cls

```
VERSION 1.0 CLASS
BEGIN
  MultiUse = -1  'True
END
Attribute VB_Name = "clsFileUtil"
Attribute VB_GlobalNameSpace = False
Attribute VB_Creatable = False
Attribute VB_PredeclaredId = False
Attribute VB_Exposed = False
Option Explicit

Private Declare Function GetOpenFileName Lib "comdlg32.dll" Alias _
      "GetOpenFileNameA" (pOPENFILENAME As OPENFILENAME) As Long

Private Type OPENFILENAME
  lStructSize As Long
  hwndOwner As Long
  hInstance As Long
  lpstrFilter As String
  lpstrCustomFilter As String
  nMaxCustFilter As Long
  nFilterIndex As Long
  lpstrFile As String
  nMaxFile As Long
  lpstrFileTitle As String
  nMaxFileTitle As Long
  lpstrInitialDir As String
  lpstrTitle As String
  FLAGS As Long
  nFileOffset As Integer
  nFileExtension As Integer
  lpstrDefExt As String
  lCustData As Long
  lpfnHook As Long
  lpTemplateName As String
```

```
End Type

Private nNumFiles As Long

Const OFN_ALLOWMULTISELECT = &H200
'Allow the user to select multiple files (Open File dialog box only).
Const OFN_CREATEPROMPT = &H2000
'Prompt if a non-existing file is chosen.
Const OFN_ENABLEHOOK = &H20
'Use the function specified by lpfnHook to process the dialog box's messages.
Const OFN_ENABLETEMPLATE = &H40
'Use the dialog box template specifed by hInstance and lpTemplateName.
Const OFN_ENABLETEMPLATEHANDLE = &H80
'Use the preloaded dialog box template specified by hInstance.
Const OFN_EXTENSIONDIFFERENT = &H400
'The function sets this flag if the user selects a file with an extension different than the one specified by lpstrDefExt.
Const OFN_FILEMUSTEXIST = &H1000
'Only allow the selection of existing files.
Const OFN_HIDEREADONLY = &H4
'Hide the Open As Read Only check box (Open File dialog box only).
Const OFN_NOCHANGEDIR = &H8
'Don 't change Windows's current directory to match the one chosen in the dialog box.
Const OFN_NODEREFERENCELINKS = &H100000
'If a shortcut file (.lnk or .pif) is chosen, return the shortcut file itself instead of the file or directory it points to.
Const OFN_NONETWORKBUTTON = &H20000
'Hide and disable the Network button in the dialog box.
Const OFN_NOREADONLYRETURN = &H8000
'The function sets this flag if the selected file is not read-only (Open File dialog box only).
Const OFN_NOVALIDATE = &H100
'Don 't check the filename for invalid characters.
Const OFN_OVERWRITEPROMPT = &H2
'Prompt the user if the chosen file already exists (Save File dialog box only).
Const OFN_PATHMUSTEXIST = &H800

Const OFN_READONLY = &H1
Const OFN_SHAREAWARE = &H4000
Const OFN_SHOWHELP = &H10

Private Declare Function GetSaveFileName Lib "comdlg32.dll" Alias "GetSaveFileNameA" (pOPENFILENAME As
OPENFILENAME) As Long
Private Declare Function FindFirstFile Lib "kernel32" Alias "FindFirstFileA" (ByVal lpFileName As String, lpFindFileData As
WIN32_FIND_DATA) As Long
Private Declare Function FindNextFile Lib "kernel32.dll" Alias "FindNextFileA" (ByVal hFindFile As Long, lpFindFileData As
WIN32_FIND_DATA) As Long
Private Declare Function FindClose Lib "kernel32.dll" (ByVal hFindFile As Long) As Long
Private Declare Function GetShortPathName Lib "kernel32" Alias "GetShortPathNameA" (ByVal lpszLongPath As String, ByVal
lpszShortPath As String, ByVal cchBuffer As Long) As Long
Private Declare Function SetEnvironmentVariable Lib "kernel32" Alias "SetEnvironmentVariableA" (ByVal lpName As String, ByVal
lpValue As String) As Long
Private Declare Function EnableWindow Lib "user32" (ByVal hwnd As Long, ByVal fEnable As Long) As Long
Private Declare Function GetEnvironmentVariable Lib "kernel32" Alias "GetEnvironmentVariableA" (ByVal lpName As String, ByVal
lpBuffer As String, ByVal nSize As Long) As Long
Private Declare Function GetTempPath Lib "kernel32" Alias "GetTempPathA" (ByVal nBufferLength As Long, ByVal lpBuffer As
String) As Long
Private Declare Function LockWindowUpdate Lib "user32" (ByVal hwndLock As Long) As Long
Private Declare Function GetWindowsDirectory Lib "kernel32" Alias "GetWindowsDirectoryA" (ByVal lpBuffer As String, ByVal nSize
As Long) As Long
Private Declare Function GetSystemDirectory Lib "kernel32" Alias "GetSystemDirectoryA" (ByVal lpBuffer As String, ByVal nSize As
Long) As Long

Private Type FILETIME
  dwLowDateTime As Long
  dwHighDateTime As Long
End Type

Private Type WIN32_FIND_DATA
  dwFileAttributes As Long
  ftCreationTime As FILETIME
  ftLastAccessTime As FILETIME
  ftLastWriteTime As FILETIME
```

```
  nFileSizeHigh As Long
  nFileSizeLow As Long
  dwReserved0 As Long
  dwReserved1 As Long
  cFileName As String * 260
  cAlternate As String * 14
End Type

Public Function FindAllFilesInPath(sPath As String, sFilter As String, ByRef sReturnArray() As String, Optional bIncludeSubfolders
As Boolean = False, Optional bReturnPathOfFile As Boolean = False) As Long

        On Error GoTo FindAllFilesInPath_ERR

          nNumFiles = 0
          FindAllFilesInPath = privFindAllFilesInPath(sPath, sFilter, sReturnArray(), bIncludeSubfolders, bReturnPathOfFile)
          nNumFiles = 0

          Exit Function

        FindAllFilesInPath_ERR:
          Debug.Print "FindAllFilesInPath_ERR " & Err.Description
          Debug.Assert 0

End Function

Private Function FileName(ByVal sFilePath As String, Optional bNoExtension As Boolean) As String
'returns the directory that the file in the path resides in:
'ie . returns "Temp.dbf" from "c:\temp\table.dbf"
        Dim i As Integer
        Dim s As String
        Dim iBeg As Integer
        Dim sName As String
        On Error GoTo GetFileName_ERR
          sFilePath = RemoveQuotes(sFilePath)
          For i = Len(sFilePath) To 1 Step -1
            s = Mid(sFilePath, i, 1)
          '   bail when when you get first backslash (s="\"):
            If s = "\" Then Exit For
          Next

          iBeg = i + 1

          If iBeg - 1 = Len(sFilePath) Then 'is a root dir
            sName = Left(sFilePath, 1)
          Else
            sName = Mid(sFilePath, iBeg)
          End If

          If bNoExtension Then
            If Len(sName) > 3 Then
          '   If there is an extension:
              If Mid(sName, Len(sName) - 3, 1) = "." Then
                If Len(sName) > 4 Then
                  FileName = Mid(sName, 1, Len(sName) - 4)
                Else
                  FileName = ""
                End If
              Else
                FileName = sName
              End If
            Else
          '   certainly no extension- filename is only 3 characters (ie. pat)
              FileName = sName
            End If
          Else
            FileName = sName
          End If

          Exit Function
```

```
        GetFileName_ERR:
            Debug.Assert 0
            Debug.Print "GetFileName_ERR: " & Err.Description
        '   return the last thing we got:
            If Len(sName) > 0 Then
                FileName = sName
            Else
        '   else pass back what we got in:
                FileName = sFilePath
            End If
End Function

Private Function RemoveQuotes(ByVal sTempExp) As String
        '   if there are quotes or double quotes around the passed in expression,
        '   this function returns the expression without them.....
        On Error GoTo RemoveQuotes_ERR
            sTempExp = Trim(sTempExp)
            RemoveQuotes = sTempExp
            If sTempExp <> Chr(34) And sTempExp <> Chr(39) Then
                If (Left(Trim(sTempExp), 1) = Chr(34) Or Left(Trim(sTempExp), 1) = Chr(39)) _
                Or (Right(Trim(sTempExp), 1) = Chr(34) Or Right(Trim(sTempExp), 1) = Chr(39)) Then

                    If Left(Trim(sTempExp), 1) = Chr(34) Or Left(Trim(sTempExp), 1) = Chr(39) Then
                        sTempExp = Mid(sTempExp, 2)
                    End If

                    If Right(Trim(sTempExp), 1) = Chr(34) Or Right(Trim(sTempExp), 1) = Chr(39) Then
                        sTempExp = Mid(sTempExp, 1, Len(sTempExp) - 1)
                    End If

                    RemoveQuotes = sTempExp
                Else
                    RemoveQuotes = sTempExp
                End If
            Else
                RemoveQuotes = ""
            End If
            Exit Function
        RemoveQuotes_ERR:
            Debug.Assert 0
            Debug.Print "RemoveQuotes_ERR: " & Err.Description
End Function

Public Function GetFileToOpen(Optional sDefaultPath As String, Optional sFilter As String, Optional lOwnerHwnd, Optional
bMultiSelect As Boolean) As String
        Dim OpenFile As OPENFILENAME
        Dim lReturn As Long
        Dim s As String
        Dim sInitFile As String
        Dim sInitDir As String

        On Error GoTo GetFileToOpen_ERR

            If Len(sDefaultPath) > 1 Then
                sInitDir = sDefaultPath
            End If
            OpenFile.lStructSize = Len(OpenFile)
            If Not IsMissing(lOwnerHwnd) Then
                OpenFile.hwndOwner = CLng(lOwnerHwnd)
            End If

            OpenFile.hInstance = App.hInstance
            OpenFile.lpstrFilter = sFilter
            OpenFile.nFilterIndex = 1
            OpenFile.lpstrFile = String(257, 0)
            OpenFile.nMaxFile = Len(OpenFile.lpstrFile) - 1
            OpenFile.lpstrFileTitle = OpenFile.lpstrFile
            OpenFile.nMaxFileTitle = OpenFile.nMaxFile
            If Len(sInitDir) > 1 Then
                OpenFile.lpstrInitialDir = sInitDir
```

```
            Else
               OpenFile.lpstrInitialDir = CurDir()
            End If

            OpenFile.lpstrTitle = ""

            OpenFile.FLAGS = OFN_FILEMUSTEXIST
            If bMultiSelect Then OpenFile.FLAGS = OpenFile.FLAGS Or OFN_ALLOWMULTISELECT

            lReturn = GetOpenFileName(OpenFile)

            If lReturn = 0 Then
               GetFileToOpen = ""
            Else
               s = OpenFile.lpstrFile
               ReturnWOEndNull s
               GetFileToOpen = s
            End If
            Exit Function
         GetFileToOpen_ERR:
            Debug.Assert 0
            Debug.Print "GetFileToOpen_ERR: " & Err.Description
            Resume Next
End Function

Private Sub ReturnWOEndNull(ByRef sString As String)
         Dim i As Integer
         Dim s As String
         '   return a string up to its terminating null character
         On Error GoTo ReturnWOEndNull_ERR
            For i = 1 To Len(sString)
               s = Mid(sString, i, 1)
               If s = Chr(0) Then
                  Exit For
               End If
            Next

            sString = Mid(sString, 1, i - 1)

            Exit Sub

         ReturnWOEndNull_ERR:
            Debug.Assert 0
            Debug.Print "ReturnWOEndNull_ERR: " & Err.Description

End Sub

Public Function GetFileToSave(Optional sDefaultPath As String, Optional sDefExtention As String, Optional sFilter As String,
Optional lOwnerHwnd) As String
         Dim OpenFile As OPENFILENAME
         Dim lReturn As Long
         Dim s As String
         Dim sInitFile As String
         Dim sInitDir As String

         On Error GoTo GetPathToSave_ERR

            If Len(sDefaultPath) > 1 Then
               sInitDir = modStringHelper.FileDir(sInitFile)
            End If

            OpenFile.lStructSize = Len(OpenFile)
            If Not IsMissing(lOwnerHwnd) Then
               OpenFile.hwndOwner = CLng(lOwnerHwnd)
            End If

            OpenFile.hInstance = App.hInstance
            OpenFile.lpstrFilter = sFilter
            OpenFile.nFilterIndex = 1
            OpenFile.lpstrFile = String(257, 0)
```

```
            OpenFile.nMaxFile = Len(OpenFile.lpstrFile) - 1
            OpenFile.lpstrFileTitle = OpenFile.lpstrFile
            OpenFile.nMaxFileTitle = OpenFile.nMaxFile

            If Len(sInitDir) > 1 Then
               OpenFile.lpstrInitialDir = sInitDir
            Else
               OpenFile.lpstrInitialDir = CurDir()
            End If

            OpenFile.lpstrTitle = ""

            If Len(sDefExtention) > 0 Then
               OpenFile.lpstrDefExt = sDefExtention
            End If

            OpenFile.FLAGS = OFN_OVERWRITEPROMPT
            lReturn = GetSaveFileName(OpenFile)
            If lReturn = 0 Then
               GetFileToSave = ""
            Else
               s = OpenFile.lpstrFile
               ReturnWOEndNull s
               GetFileToSave = s
            End If

            Exit Function

         GetPathToSave_ERR:
            Debug.Assert 0
            Debug.Print "GetPathToSave_ERR: " & Err.Description
            Resume Next
End Function

Private Function privFindAllFilesInPath(sPath As String, sFilter As String, ByRef sReturnArray() As String, Optional
bIncludeSubfolders As Boolean = False, Optional bReturnPathOfFile As Boolean = False) As Long

         ' Search for all files that match "C:\MyProgram\user*.*".  Display
         ' the filename of each file that matches the string.
         Dim hsearch As Long  ' handle to the file search
         Dim findinfo As WIN32_FIND_DATA  ' receives info about matching files
         Dim success As Long  ' will be 1 if successive searches are successful, 0 if not
         Dim buffer As String  ' string buffer to use to process the filename(s)
         Dim retval As Long  ' generic return value
         Dim n As Long
         Dim FSO As FileSystemObject
         Dim fol As Folder
         Dim subfol As Folder

         On Error GoTo FindAllFilesInPath_ERR

            If Right(sPath, 1) <> "\" Then sPath = sPath & "\"

            Set FSO = New FileSystemObject
            Set fol = FSO.GetFolder(sPath)

            ' Begin a file search:
            hsearch = FindFirstFile(sPath & sFilter, findinfo)

            If hsearch > -1 Then  ' no files match the search string

               ' Display name of each file that matches the search.  Note that the name is displayed, the
               ' next file (if any) is found, and then the loop restarts.  This way the first file
               ' (found above) will also be displayed.
               Do  ' begin loop
               ' Extract the filename from the fixed-length string:
               buffer = Left(sPath & findinfo.cFileName, InStr(sPath & findinfo.cFileName, vbNullChar) - 1)

               If Not bReturnPathOfFile Then
                  buffer = FileName(buffer)
```

```
                End If

                ReDim Preserve sReturnArray(nNumFiles)
                sReturnArray(nNumFiles) = buffer
                nNumFiles = nNumFiles + 1
                n = n + 1

                ' Get the next matching file and loop if it exists:
                success = FindNextFile(hsearch, findinfo)
                Loop Until success = 0  ' keep looping until no more matching files are found

                ' Close the file search handle
                retval = FindClose(hsearch)

            End If
            If bIncludeSubfolders Then
                For Each subfol In fol.SubFolders
                    n = n + privFindAllFilesInPath(subfol.Path, sFilter, sReturnArray)
                Next
            End If
            privFindAllFilesInPath = nNumFiles
            Exit Function

        FindAllFilesInPath_ERR:
            Debug.Print "FindAllFilesInPath_ERR: " & Err.Description
            Debug.Assert 0
End Function
```

FrmProgress.frm

```
VERSION 5.00
Begin {C62A69F0-16DC-11CE-9E98-00AA00574A4F} frmProgress
   Caption         =   "UserForm1"
   ClientHeight    =   3810
   ClientLeft      =   45
   ClientTop       =   330
   ClientWidth     =   4425
   OleObjectBlob   =   "frmProgress.frx":0000
   StartUpPosition =   1  'CenterOwner
End
Attribute VB_Name = "frmProgress"
Attribute VB_GlobalNameSpace = False
Attribute VB_Creatable = False
Attribute VB_PredeclaredId = True
Attribute VB_Exposed = False
Public blnCancelPressed As Boolean
Private mintCur As Integer
Private mintTotal As Integer
Public datStart As Date

Private Sub cmdCancel_Click()
            blnCancelPressed = True
End Sub
Public Sub UpdateProgress(intCurrent As Integer, intTotal As Integer)

            mintCur = intCurrent
            mintTotal = intTotal

            If mintTotal = 0 Then
                Exit Sub
            End If

            lblCounter.Caption = intCurrent & " of " & intTotal

            ProgressBar1.Value = intCurrent / intTotal * 100
            ProgressBar1.Refresh

            UpdateTime
            Me.Repaint
End Sub
```

```
Private Sub UserForm_Initialize()
                datStart = Now
                lblStarted.Caption = "Time Started : " & datStart

End Sub
Sub UpdateTime()
   If mintCur = 0 Then
      Exit Sub
   End If

   Dim intMinPast As Long
   Dim intSecPast As Long
   intSecPast = DateDiff("s", datStart, Now)
   intMinPast = intSecPast \ 60
   Dim strLapsed As String
   strLapsed = intMinPast & ":" & intSecPast Mod 60
   Dim lngSecsRemain As Double
   "remaining = items remaining * lapsed / completed
   lngSecsRemain = (mintTotal - mintCur) * (intSecPast / mintCur)
   Dim strRemain As String
   strRemain = lngSecsRemain \ 60 & ":" & lngSecsRemain Mod 60

   lblStarted.Caption = "Time Started : " & datStart
   lblElapsed.Caption = "Time Elapsed : " & strLapsed
   lblRemaining.Caption = "Time Remaining : " & strRemain

End Sub
```

APPENDIX Q. SAMPLE DISTANCE AND ANGLE TABLE

RowID – unique id for each distance and angle calculation
SiteID – unique 3-character ID for each air monitoring station (used as receptors in the pilot study to calibrate model) and a unique 7-digit ID for each school (used as receptors in the state model run)
GridID – unique ID for the ¼ mile model gridcell to which distance and angle are being measured (from the receptor)
Distance – distance from the receptor to the model gridcell centroid (meters)
Angle – angle between the receptor and gridcell centroid from 0 to 360, which is measured clockwise and where 0 is due north (degrees)

```
ROWID, SITEID, GRIDID, DISTANCE, ANGLE
1,ARB,1750755,42237.25,346.32
2,ARB,1750756,41988.39,348.46
3,ARB,1750757,41779.78,350.63
4,ARB,1750758,41612.01,352.83
5,ARB,1750759,41485.59,355.05
6,ARB,1750760,41400.88,357.29
7,ARB,1750761,41358.16,359.53
8,ARB,1750762,41357.54,1.78
9,ARB,1750763,41399.03,4.03
10,ARB,1750764,41482.5,6.27
11,ARB,1750765,41607.7,8.48
12,ARB,1750766,41774.26,10.68
13,ARB,1750767,41981.68,12.84
14,ARB,1750768,42229.37,14.97
15,ARB,1752084,41956.47,339.55
16,ARB,1752085,41580.42,341.61
17,ARB,1752086,41243.2,343.71
18,ARB,1752087,40945.78,345.87
19,ARB,1752088,40689.02,348.07
20,ARB,1752089,40473.71,350.3
21,ARB,1752090,40300.51,352.57
22,ARB,1752091,40169.96,354.86
23,ARB,1752092,40082.47,357.17
24,ARB,1752093,40038.34,359.5
25,ARB,1752094,40037.7,1.82
26,ARB,1752095,40080.56,4.14
27,ARB,1752096,40166.77,6.45
28,ARB,1752097,40296.06,8.74
29,ARB,1752098,40468.02,11.
30,ARB,1752099,40682.1,13.23
```

APPENDIX R. ATTACHED SPREADSHEET FILES

- *schools.location.verifylist.xls*
- *duplicates.final.xls*
- *summary.of.duplicates.final.xls*
- *corrupted _nn.xls*
- *outliers.final.xls*
- *Site.xls*