FINAL REPORT TO THE CALIFORNIA DEPARTMENT OF FOOD AND AGRICULTURE FOR CONTRACT AGREEMENT NO. 98-0241

DATA QUALITY OF CALIFORNIA'S PESTICIDE USE REPORT

By
Larry Wilhoit, Minghua Zhang, Lisa Ross

August 2001

STATE OF CALIFORNIA
Environmental Protection Agency
Department of Pesticide Regulation
Pest Management and Licensing Branch
Pest Management Analysis and Planning Program
Sacramento, California 95812-4015

PM 01-02

EXECUTIVE SUMMARY
Of Report PM 01-02 Entitled
"Final Report to the California Department of Food and Agriculture
For Contract Agreement No. 98-0241
Data Quality of California's Pesticide Use Report"

Environmental Protection Agency
Department of Pesticide Regulation
Pest Management and Licensing Branch

BACKGROUND

The California Department of Pesticide Regulation's (DPR's) Pesticide Use Report (PUR) is probably the largest and most complete database on pesticide use in the world. A system to collect information on pesticide use in California has been in operation in some form for over 50 years, with the current use reporting system started in 1990. The PUR contains information on nearly all production agricultural pesticide use and some nonagricultural use in California. The data collected include the pesticide product used, the date it was applied, the particular field treated, and application location to a square-mile section. Production agricultural use includes applications to growing crops, agricultural fields, and most applications to forest trees and ornamental turf. Other pesticide uses reported to DPR include post-harvest commodity treatments, right of way, landscapes, structural use, and other nonagricultural uses by commercial applicators.

DPR expanded pesticide use collection in 1990 primarily to more accurately assess dietary risk as well as exposure and potential risk to workers. However, the PUR is also used for a wide variety of environmental and public health purposes. To ensure the accuracy of results from these uses and any regulatory decisions based on such assessments, PUR data must adequately represent actual pesticide use in the state.

Because of the importance of the PUR for many groups and individuals, it is critical that the database be as accurate and complete as possible. People who use the data need to feel confident that the data are sufficiently accurate for their purposes. Bad data are worse than no data. If a system fails to minimize errors, two serious problems will arise. Some people will use the data not knowing about the errors and, therefore, make wrong conclusions. If laws and regulations are based on these conclusions, there is the potential for serious consequences. Other people will not trust the data and therefore ignore it. In either case, the huge effort at data collection is wasted.

PURPOSE

In an effort to minimize errors and improve the quality of the data in the PUR, DPR's Pest Management Analysis and Planning Program entered into a contract with the Office of Pesticide Consultation and Analysis of the California Department of Food and Agriculture (CDFA). Improvement in data quality was focused on four main goals: (1) identify specific PUR data fields and information in need of improvement, (2) assess the quality of spatial data in the PUR, (3) improve error-checking procedures used when loading data into the PUR database, and (4) conduct a survey of county staff to identify the variation in data-field definitions and procedures used for data entry into the PUR. In this final contract report, the first three goals are described. Goal four will be completed at a later date and results related to the acres-planted questions in the survey will be summarized for CDFA.

IDENTIFICATION OF DATA FIELDS AND OTHER INFORMATION

At the onset of the contract, staff from DPR and CDFA met to discuss the data fields most important to CDFA that needed improvement. The most critical data fields identified were acres treated, acres planted, and rate of pesticide use. These needs were incorporated into the contract work as described in more detail below.

In addition to meeting with CDFA staff, DPR also established a committee of various PUR users to identify their concerns about the PUR database as it impacts their assessments. This committee met from September 1999 to May 2000 and discussed many issues related to data quality and produced a list of issues they hoped would be addressed by DPR (Appendix A). This committee also helped organize and plan a conference on the PUR from which an expanded issues list was developed (Appendix A). The conference drew over 200 attendees who listened to seminars on various uses for PUR data. These uses included topics such as PUR data role in policy and regulatory decision-making, human health and environmental quality assessments, exposure and epidemiological studies, economic analyses, and integrated pest management studies. In addition, attendees participated in concurrent sessions designed to reveal problems encountered when using PUR data. Attendance at the conference demonstrated the importance of the PUR database to a wide variety of people and organizations. It also illustrated the need for accurate PUR data given the various links made between pesticide use, environmental impacts, human health assessments, and policy/regulatory decisions.

In response to the conference and the list of issues developed over the prior months, DPR staff developed a work plan for addressing these issues (Appendix B). Updates to the work plan can also be found on DPR's web page at www.cdpr.ca.gov under "Programs and Services". In the coming months and years, DPR hopes to address these various issues.

ASSESSMENT OF DATA QUALITY OF SPATIAL ATTRIBUTES IN THE PUR

Various data fields in the PUR database were assessed for their potential rate of error from 1990 through 1997 (Appendix C). Most of the data fields assessed relate to spatial attributes of pesticide applications. These include county code, meridian/township/range/section (MTRS) designation, grower identification number and site location identification. Also assessed were the data fields for crop planted, area planted, and area treated. In addition, agricultural records (rows in the PUR database pertaining to agricultural use) were checked for potential duplicates. Errors in these data fields could lead to erroneous assessments about the amount of pesticide used and location of use relative to sensitive sites. Errors in these types of data fields could have consequences for community, environmental, and worker health assessments as well as for the regulated community. Therefore, it is important to understand the rate of error found in these data fields so the data can be used with a certain degree of confidence. It is also important when planning for future improvements to the PUR database.

Statewide, potential error rates averaged over the years 1990 to 1997 were about 5% or less of all agricultural records collected (Appendix C). One exception was the potential error rate calculated for acres planted. The statewide error rate averaged 8.1% of all agricultural fields and 17% of all agricultural records collected from the years 1990 to 1997. The specific reason for such a relatively high error rate is not certain but could relate to the inconsistent use of site location identifiers that should uniquely identify each agricultural field.

Examination of potential error rates over the eight-year period statewide and by county, indicated variation temporally and spatially. Three of seven error types indicate a decline in the rate of error averaged statewide. However, duplicate records, MTRS inconsistency, acres planted, and acres treated did not show a similar trend. By county, error rates typically declined over the years with some exceptions. Error-rate details by county and year are provided in Appendix C. The accuracy of any environmental or human health assessment will depend in part on the accuracy of the pesticide use data used for those assessments. Therefore, it is important to be aware of the error rates encountered in the PUR database, and any details on the data field, county, and year, as applicable to the assessment. Appendix C provides those details and could be used by researchers to help assess the degree of accuracy of their work.

THE PUR LOADER AND ERROR HANDLING PROCESSES

In concert with the assessment of error rates in selected data fields of the PUR (Appendix C), the loader program was modified to reduce the number of errors in the database. In addition, DPR recently decided to convert its PUR database from FOCUS to Oracle. Since this involved rewriting the program to load and error check the data, this conversion provided an opportunity to improve the error-checking procedures. We were able to add several new error checks and improve some of the previous checks thanks to the monetary support from CDFA.

The PUR loader program is a computer program that loads data from pesticide use reports into an Oracle database (Appendix D). These data are first entered into databases at each County Agricultural Commissioner (CAC) office and then sent to DPR where the loader program is run. In addition to loading the data, the program also searches for errors, records any errors found, and corrects errors where possible.

During collection and processing of the PUR data, several kinds of error checks occur, both at the CAC offices where the data are entered and during the loading of the data from the county into DPR's PUR database. Ideally, error checking should occur at the time the data are first entered, but until the data collection system is changed, DPR can make improvements to the data received from the county.

The improved loader program carries out several steps (see Appendix D for details):

1. Checks that data from each PUR county data file sent to DPR has not already been loaded.
2. Checks that the new data file has the correct structure and fixes certain kinds of errors.
3. Loads all the data from the new file directly into an Oracle table.
4. Logs the name and date of the county file, number of rows of data entered, and the date loaded.
5. Checks that each record in the new file is not an erroneous duplicate of records already loaded.
6. Checks each data field in this table for a series of possible errors.
7. Corrects errors or makes estimates to replace invalid data where possible, otherwise replaces the value with a blank or leaves it unchanged.
8. Records any uncorrected errors it finds including both the original value and new value, and the kind of error.
9. Records any changes made to the data, the date of the change, and whether it was corrected, estimated, or replaced with a null.
10. Identifies each agricultural field, assigns it a field identification code, determines the most likely acres planted and location of the field, and creates a record of this agricultural field.
11. If the agricultural field has records in the PUR with different values reported for its acres planted or location, the agricultural field is marked as inconsistent and a list of all inconsistent values of acres planted and location is made for this field.
12. Makes some calculations and conversions (such as getting the DPR product identification number from the pesticide registration number and the pounds of active ingredient used from the amount of product used).
13. Loads the valid and converted data into another table.
14. Creates and prints an error report with all the errors found in the PUR data; and
15. E-mails loader log files to the loader administrators; these report the data files successfully loaded, the files not loaded because of errors, and any database or operating system error messages that may have been produced during loading.

The new loader program includes several error-checking routines that were not previously performed. These error checks are listed below and described in more detail in Appendix D:

1. Product identification number.
2. Grower identification number.
3. Inconsistent values for an agricultural field.
4. Duplicate records.
5. High rates of use.

Once various errors are identified and stored in a series of data files, the counties receive a file containing the records in need of data improvement. Staff at the CAC office correct the records, wherever possible, and return the data to DPR. An error-correction program will be developed to enter corrected data into the PUR and will be used for the first time on the 1999 PUR preliminary data. Once this database is corrected, we plan to compare the rate of errors found in 1999 with prior years. This will help us identify where to focus future data improvement efforts.

CONCLUSIONS

Much time, effort, and resources are spent on collecting pesticide use data in California because of the need for such information. Many groups and individuals use PUR data for various human health and environmental assessments. These assessments may lead to policy and/or regulatory decisions that affect the agricultural community as well as others who use pesticides. Because of the importance of the PUR database, it is critical that these data be as accurate and complete as possible.

In our efforts here, we have made an initial attempt to assess the rate of error in certain data fields in the database as well as established an error-checking process to catch as many errors as possible. Although error rates in certain data fields have declined since 1990, these measures are by no means complete and efforts to improve data quality of the PUR are on going. Future efforts, in addition to those outlined in DPR's PUR Improvement Plan, will include assessing improvements made with our new PUR Loader Program and evaluating a county survey used to identify variation in use and definition of data fields in the PUR.

TABLE OF CONTENTS

Page

APPENDIX A

Preliminary and Expanded Lists of PUR Problems and Issues

Preliminary List of PUR Problems and Issues
For Discussion at the PUR Conference, May 8, 2000

Purposes:

1. Information for DPR management and conference attendees.
2. To facilitate discussion by conference attendees of additional problems and issues faced by PUR users.
3. To provide an outline for the PUR Technical Advisory Committee's report to DPR on problems with the PUR that affect work done by committee members

Outline of Problems and Issues

1. **Data Collection**
   A. PUR regulations and requirements
      i. Unclear regulations. Current regulations are confusing, incomplete, or result in inaccurate values. These confusions sometimes result in errors by applicators in filling out their use reports.
      ii. Enforcement letters are not unified into one document. Some of the procedures were more fully explained in a series of enforcement letters that were distributed to the counties after the regulations were instituted. Confusion on data entry protocol has resulted because the enforcement letters were never organized clearly into a single document (manual).
      iii. Above issues lead to inconsistency among counties particularly for items B and C below. This makes it difficult to understand the data since the meaning of the data may vary for different counties.
   B. General crop category vs. specific crop. Sometimes a general crop category is entered into the PUR database instead of the specific crop grown. For example, a grower may record a crop as "citrus" rather than "oranges" because the pesticide is labeled for use on "citrus." For many of our analyses we need to know the specific crop, not a general crop category.
      i. Qualifier code. The qualify_code, which can be used to provide more specific crop information, is usually not used. Unfortunately, the few counties that do use the qualify_code do not follow the same protocol for determining its value.
      ii. Planting sequence. The PUR data field planting_seq is usually not reported for crops that have multiple plantings in a year. There are many instances where analysts need to know how many plantings were done, the timing of the plantings, or how much pesticide was used per planting.
   C. Site location identification. Counties are inconsistent in definition of site_loc_id and in some cases do not clearly designate a geographical location. This information is important when we need to make calculations on field level statistics. It is also necessary to arrive at total acres planted and % crop treated.

D. Acres treated are sometimes reported as the entire field even when only part of the field is treated. This is apparently the recommended practice given in an enforcement letter for strip treatments. However, this practice produces incorrect statistics for acres treated, overestimates percent acres treated, and underestimates the rate of use.
E. The data source for acres planted is variable. Data for this field are transferred from the grower's permit or from the PUR reporting form. If taken from the permit, this field is not always updated when the actual acres planted (as indicated on the PUR reporting form) is different. This field has been used to determine the number of acres planted of each crop in California. However, for some crops the acres planted from the PUR differs considerably from the County Agricultural Commissioner's (CAC) crop reports and the California Department of Food and Agriculture (CDFA) values. Why do these different sources differ so much and which values are the more accurate?
F. The number of applications is not consistently reported in monthly summary reports. Thus, it is difficult or impossible to interpret these data.
G. Reporting amount, area, or volume treated in commodity fumigation. People will sometimes use or report fumigation of an entire warehouse even if only one crate is actually treated.
H. Amount of product used. Growers sometimes will report the total amount of pesticide plus diluent used, *i.e.*, the total volume of material in the tank; they should report the amount of the pesticide product only. This results in too much product reported.

2. Data quality

A. Data validation at CAC offices is inadequate. The county data entry programs do not check for all of the errors that are checked for at DPR.
B. Poor data entry screen design. Errors in typing are difficult for data entry staff to check, often owing to poor data entry design. For example, units of measure are coded as numbers rather than the actual name of the units.
C. Illegible PUR forms.
D. Duplicate submissions or data entry. Sometimes records are entered twice or sent twice to DPR, resulting in duplicate records. Some records seem to have been "fixed," but the original record is not always deleted. This results in over reporting applications and amounts applied.
E. Lack of error corrections. Counties do not always correct errors that are sent back to them.

3. **Data Access**
   A. Data formats. Some of the data formats used in the files that are distributed are difficult to use by some people, for example, the fixed field text files on the PUR CD-ROM.
   B. Incomplete documentation. Users need access to a discussion of critical data analysis issues in order to accurately handle PUR data. Some of these problems were outlined in section 1.A, for example, the differences between counties in how some data fields are defined or used.
   C. Limited summary statistical data available. There is no easy method to view different kinds of summary statistics from the PUR except that provided in a limited set of PUR summary tables.
   D. Active ingredient (AI) names. Some of AI names and classifications are inconsistent with other databases or contain errors.
      i. Different databases (DPR, U.S. Environmental Protection Agency (US EPA), U.S. Department of Agriculture (USDA)) use different AI names making it difficult to compare databases. These institutions have chemical synonym databases that should be synchronized and users need to be informed of potential naming discrepancies.
      ii. DPR produces reports with very specific chemical classifications of some AIs but most researchers are interested in a more broadly defined AI classification. For example, DPR lists more than 20 different esters and salts of 2,4-D but in summary reports, an aggregate summation for all 2,4-D variants would be more useful.
      iii. The oils are particularly troublesome; oils are listed under several different names that are not clearly explained and appear to be inconsistent.
   E. Crop names. DPR uses different crop names and classifications than other agencies (such as USDA and CDFA), making it difficult to compare statistics.
   F. Adjuvants and inert ingredients. Adjuvants are not clearly and accurately separated from other AIs. Some products are registered as adjuvants but the Label Database is inaccurate in recording this. In other cases, an adjuvant may be considered as part of a product and would not be separately reported. Thus, information on adjuvant use in the PUR is incomplete and can't be interpreted.
   G. Precision of data. Some data are reported to 4 decimal places when measurement of the particular parameter is less refined.

4. **Additional data needs**
   A. Target pest. Lack of pest information makes it difficult to monitor pest infestations or target IPM programs. It also misses an opportunity to remind growers frequently about the importance of pest identification (a basic principle of IPM). Alternative pest control analyses (including economic impact analysis of regulatory issues) would be considerably more defensible if analysts knew how pests were being controlled from PUR data.
   B. Date crop planted. This information would help users distinguish separate plantings. One example where this is important is determining the total acres planted of crops with a growing season that spans two calendar years. The acres planted, calculated from the PUR for these crops, will be overestimated when pesticide use is reported for that crop in two different years. We currently have no way of knowing whether a field had two different plantings or not. Since most crops do not span two years, we typically assume fields are different from one year to the next.
   C. Date crop harvested. This information is used by US EPA and DPR's medical toxicology unit in their residue analysis programs.
   D. Acres harvested. This would help users relate use to what gets into the market.
   E. Crop yield
   F. Additional site information
      i. Schools, day-care centers
      ii. Federal land
      iii. Rights of way
      iv. Water bodies
      v. Indian tribal lands
      vi. Homeowners
      vii. Farm animals
      viii. Institutions
      ix. Organic fields
      x. Genetically modified crops
      xi. Hospitals
      xii. Proximity to sensitive sites, e.g. schools, organic fields, etc.
   G. Identify cancellations of chemical or chemical/commodity combinations.
   H. Label rate.
   I. Pesticide type (insecticide, herbicide, etc) or class (OP, carbamate, etc). DPR's Label Database provides this information, but the information in this database have is inaccurate and the categories are not necessarily easy to apply.
   J. Identify section 18 applications.

5. **Miscellaneous issues**
   A. Ongoing compliance monitoring. Users of the data need to know the level of reporting compliance to have a sense of the completeness of the data.
   B. Errors in the Sales database. The Sales database has errors. It is used to get a general idea of possible pesticide use for which there are no reporting requirements and to get estimates of the level of reporting compliance.
   C. Label Database.
      i. Errors. The label database contains errors that affect data in the PUR, such as incorrect percent AI or specific gravity in pesticide products and errors in the classification of AIs.
      ii. Crop names on labels. The PUR loses information on crop specificity because DPR's error checking process requires the applicator report only names of crops listed on the label. For example, if an application occurs to navel oranges and the label only lists citrus, citrus will be reported as the crop treated, not navel oranges.
      iii. Maximum label rates. The label database does not contain label rates. This information is critical for error checking.
   D. PUR error reporting and correction procedures.
      i. Procedure for reporting errors. There is no procedure for reporting to DPR errors found in the PUR (by people in and outside of DPR).
      ii. Reporting errors found and corrections made to the public. DPR has no process for informing PUR data users about recently found errors or corrections made to the database.
      iii. Peer review of DPR error-checking procedures. DPR has not had their error-checking procedures reviewed by knowledgeable people from outside of DPR.

EXPANDED LIST OF PUR PROBLEMS AND ISSUES

Background: This list was formulated as a result of discussions with a technical advisory committee convened by DPR, as well as issues raised at the PUR conference.

1. DATA COLLECTION ISSUES

A. PUR regulations and requirements

1. Unclear regulations
2. Enforcement letters not unified into one document
3. Above issues lead to inconsistency among counties particularly for items B and C below
4. Do not allow counties to re-use retired grower_ids

B. General crop category vs. specific crop

1. Qualifier codes
2. Planting sequence

C. Site location ID

1. Require all reports of different applications to one field in a year to use the same grower_id [Editor: site_loc_id?] each time, so all fields can be uniquely identified).

D. Acres treated

1. Spot treatments
2. Strip/band treatments

E. Acres planted

F. Number of applications not consistently reported in monthly summary reports

G. Reporting amount, area, or volume treated in commodity fumigation

H. Amount of product used: Product vs. diluted amount.

I. PLSS reporting errors. Consider hiring an outside contractor to conduct call backs to verify information, similar to the reliability checks that epidemiologists perform.

2. DATA QUALITY ISSUES

A. Data validation at CAC offices (or field level)

B. Poor data entry screen design

C. Illegible PUR forms

D. Duplicate submissions or data entry

E. Lack of error corrections

F. In the amount used field in the PUR convert all values to either pounds or gallons (don't use measures such as pints, quarts, liters, etc.)

3. DATA ACCESS ISSUES

A. Data formats:
1. Fixed field file structure on CD-ROM is a problem for some users.
2. Provide data as an Oracle dump file.

B. Incomplete technical documentation

1. Data dictionary with informational notes
2. Example SQL and PL/SQL code

C. Limited summary statistical data available
1. Map on watershed basis using GIS.
2. Be able to sort lists on the Web site.

D. Active ingredient names

1. Inconsistent names in various databases (DPR, US EPA, USDA)
2. Specific vs. more general names (eg 2,4-D vs list of all the different esters and salts of 2,4-D)
3. Categories for the different kinds of oils
4. Include trade names

E. Crop names. Inconsistent names in various databases

F. Adjuvant and inert ingredient reporting

G. Inappropriate precision in reports

H. Data Integration. Link the PUR with other databases (e.g. DPR's illness reporting system).

1. Provide better links in the data set to information needed for modeling, e.g., environmental half lives.
2. Connect use and permit data.
3. Identify toxicity category (e.g. carcinogens, etc.)
4. Relate to residue monitoring data at the field level.
5. Relate sales data to use data.

I. Allow growers to view a history of their past use (this might be a good incentive for them to use a Web-based reporting system).

J. Release PUR data monthly

K. Do not eliminate all outliers.

4. ADDITIONAL DATA NEEDS

A. Target pests
1. Categorize target pests

B. Date crop planted

C. Date crop harvested

D. Acres harvested

E. Crop yield

F. Additional site information (change monthly summary to site specific)

1. Schools, day-care centers
2. Federal land
3. Rights of way
4. Water bodies
5. Indian tribal lands
6. Homeowners
7. Farm animals
8. Institutions
9. Organic fields
10. GM crops
11. Hospitals

12. Proximity to sensitive sites, e.g. schools, organic fields, etc.
13. Senior care centers
14. Golf courses

G. Identify cancellations of chemical or chemical/commodity combinations

H. Label rate

I. Pesticide type (insecticide, herbicide, etc) and class (OP, carbamate, etc)

J. Identify section 18 applications

K. Geography. Link PUR data to satellite imagery, aerial photography, and land use survey data sets.

L. Identify applicator (PCO) that made the application, and PCA affiliation.

M. Collect pesticide sales data by county.

N. Collect more information on structural and industrial use

O. Identify genetically modified organisms (GMOs)

5. MISCELLANEOUS ISSUES

A. Ongoing compliance monitoring

B. Errors in sales database

C. Label Database

1. Errors in percent AI and specific gravity

2. Crop names on labels

3. Maximum label rates

D. PUR error reporting and correction procedures
1. Procedure for reporting errors
2. Reporting errors found and corrections made to public
3. Peer review of DPR error-checking procedures
4. Find incentives to promote error correction.

E. Link use data with environmental media data (i.e. not only where the pesticide was released, but where it goes).

F. Notification. Use real-time reporting methods, e.g. using GIS systems, to provide advance warning of an application. Useful for schools, daycare centers, etc.

G. Other data sources. For example, what data elements in 'Doanes' sources are considered proprietary, and what makes it so valuable to US EPA?

APPENDIX B

Pesticide Use Report Improvement Plan

PESTICIDE USE REPORT IMPROVEMENT PLAN

Issue	Improvement Requirements	Status/Plan
I. Data Collection		
A. PUR regulations and requirements		
1. Unclear regulations	2	
2. Enforcement letters not unified into one document	1	
3. Inconsistency among counties	1	EMPM conducted a survey of county practices
4. Counties can re-use grower_ids	1	
General crop category vs. specific crop		
1. Qualifier codes	1, 5, 6	Commodity Code Workgroup will review & make recommendation.
2. Planting sequence	1, 5, 6	Commodity Code Workgroup will review & make recommendation.
C. Site location ID		
1. Require all reports of different applications to one field in a year to use the same site_loc_id	1	
D. Acres treated		
1. Spot treatments	1, 5	
2. Strip/band treatments	1, 5	
E. Acres planted	1, 6	
F. Number of applications not consistently reported in the monthly summary	1	
G. Reporting amount, area, or volume treated in commodity fumigation	1, 5	
H. Amount of product used: product vs. diluted amount	1, 4, 5	Rate/acre validation expanded for 1998 PUR.
I. PLSS reporting errors	7	Additional error check added for 1999 PUR.
II. Data Quality		
A. Data validation at CAC offices (or field level)	7	
B. Poor data entry screen design	7	Concept accepted to review the permit/PUR program.
C. Illegible PUR forms	7	Fillable forms on external web 9/2000.
D. Duplicate submissions or data entry	7	Additional error check added for 2000 PUR.
E. Lack of error corrections	7	Correction rate for 1999 PUR currently at 99.45%.
F. Convert amount used to either pounds or gallons	7	
III. Data Access		
A. Data formats		
1. Fixed field file structure on CD-Rom can be problematic	7	
2. Provide data as an Oracle dump file	7	

Improvement Requirements:
1. New/amend policy
2. New/amend regulation
3. New/amend legislation
4. Change/expand label database
5. Change/expand the PUR program
6. Change/expand the permit program
7. Change/add technology

Issue	Improvement Requirements	Status/Plan
B. Incomplete technical documentation		
1. Data dictionary with informational notes	7	
2. Example SQL and PL/SQL code	7	EMPM currently in process of developing standard queries.
C. Limited summary statistical data available		
1. Map on watershed basis using GIS	7	Proposed under the Integrated Pest. Resource Directory E-Government project.
2. Be able to sort lists on the Web site	7	Proposed under the Integrated Pest. Resource Directory E-Government project.
D. Active ingredient names		
1. Inconsistent names in various databases (DPR, US/EPA, USDA)	4	AI Workgroup to be formed.
2. Specific v. more general names (e.g. 2,3-D vs. list of all the different esters and salts of 2,4-D)	4, 7	AI Workgroup to be formed.
3. Categories for the different kinds of oils	4	AI Workgroup to be formed.
4. Include trade names		AI Workgroup to be formed.
E. Crop names: inconsistent names in various databases		Commodity Code Workgroup will review & make recommendation.
F. Adjuvant and inert ingredient reporting		
G. Inappropriate precision in reports		
H. Data integration: link the PUR with other databases (e.g., DPR's Illness Reporting System)		
1. Provide better links in the data set to information needed for modeling, e.g., environmental half lives	7	Proposed under the Integrated Pest. Resource Directory E-Government project.
2. Connect use and permit data	1, 7	
3. Identify toxicity category (e.g., carcinogens, etc.)	4, 7	
4. Relate to residue monitoring data at the field level	1, 7	
5. Relate sales data to use data	7	
I. Allow growers to view a history of their past use	7	
J. Release PUR data monthly	1	
K. Do not eliminate all outliers	1	
IV. Additional Data Needs		
A. Target pests		
1. Categorize target pests	1, 2, 3, 5, 7	
B. Date crop planted	1, 2, 3, 5, 7	
C. Date crop harvested	1, 2, 3, 5, 7	
D. Acres harvested	1, 2, 3, 5, 7	
E. Crop yield	1, 2, 3, 5, 7	

Improvement Requirements:
1. New/amend policy
2. New/amend regulation
3. New/amend legislation
4. Change/expand label database
5. Change/expand the PUR program
6. Change/expand the permit program
7. Change/add technology

Issue	Improvement Requirements	Status/Plan
F. Additional site information (change monthly reporting to site specific)		
1. Schools, day-care centers	1, 2, 3, 5, 7	AB 2260 requires commercial applications to be reported at public schools and public day care facilities beginning 1/1/02.
2. Federal land	1, 2, 3, 5, 7	
3. Rights-of-way	1, 2, 3, 5, 7	
4. Water bodies	1, 2, 3, 5, 7	
5. Indian tribal lands	1, 2, 3, 5, 7	
6. Homeowners	1, 2, 3, 5, 7	
7. Farm animals	1, 2, 3, 5, 7	
8. Institutions	1, 2, 3, 5, 7	
9. Organic fields	1, 2, 3, 5, 7	
10. GM crops	1, 2, 3, 5, 7	
11. Hospitals	1, 2, 3, 5, 7	
12. Proximity to sensitive sites, e.g., schools, organic fields, etc.	1, 2, 3, 5, 7	
13. Senior care centers	1, 2, 3, 5, 7	
14. Golf courses	1, 2, 3, 5, 7	
G. Identify cancellations of chemical or chemical/commodity combinations	4, 5, 7	
H. Label rate	4, 5, 7	
I. Pesticide type (insecticide, herbicide, etc. and class (OP, carbamate, etc.)	4, 5, 7	
J. Identify section 18 applications	4, 5, 7	
K. Geography: link PUR data to satellite imagery, aerial photography, and land use survey data sets	7	Currently underway at CAC offices & DPR on a limited basis.
L. Identify applicator (PCO) that made the application and PCA affiliation	1, 5	
M. Collect pesticide sales data by county	1, 2, 3, 7	
N. Collect more information on structural and industrial uses	1, 2, 3, 5, 7	
O. Identify genetically modified organisms (GMOs)	1, 2, 3, 5	
V. Miscellaneous		
A. Ongoing compliance monitoring	1	
B. Errors in sales database	1, 7	
C. Label database		
1. Errors in percent AI and specific gravity	1, 7	Requires quality control of the label database.
2. Crop names on labels	1, 7	PUR Workgroup will address and make a recommendation.
3. Maximum label rates	4, 7	
D. PUR error reporting and correction procedures		
1. Procedure for reporting errors	1	Under development by PUR Workgroup.

Improvement Requirements:
1. New/amend policy
2. New/amend regulation
3. New/amend legislation
4. Change/expand label database
5. Change/expand the PUR program
6. Change/expand the permit program
7. Change/add technology

Issue		Improvement Requirements	Status/Plan
	2. Reporting errors found and corrections made to public	1	
	3. Peer review of DPR error-checking procedures	1	
	4. Find incentives to promote error correction	1	
E.	Link use data with environmental media data	1, 7	Is being addressed by DPR's Cal/PIP project
F.	Notification. Use real time reporting methods (e.g., schools)	1, 2, 3, 7	
G.	Other data sources	7	

Improvement Requirements:
1. New/amend policy
2. New/amend regulation
3. New/amend legislation
4. Change/expand label database
5. Change/expand the PUR program
6. Change/expand the permit program
7. Change/add technology

APPENDIX C

California's Pesticide Use Report: An Assessment of Spatial Data Quality

California's Pesticide Use Report: An Assessment of Spatial Data Quality

Pest Management Grant Final Report
Contract #: 99-0166

July 3, 2000

Minghua Zhang, Assistant Adjunct Professor, Department of Land, Air and Water Resources, University of California, Davis, CA 95616
Larry Wilhoit, Senior Research Scientist, California Department of Pesticide Regulation, Sacramento, CA 95814
John Steggall, Senior Research Scientist, California Department of Food and Agriculture, Sacramento, CA 95814

Prepared for the California Department of Pesticide Regulation

Table of Contents

Summary

The pesticide use report (PUR) database has been identified as one of the best databases for tracking pesticide applications temporally and spatially. Over two to three and a half million records have been reported annually for all pesticides applied in California and millions of dollars have been spent in the data collection process. Since the PUR is unique in tracking pesticide use, researchers, environmentalists, and regulators nationally and internationally have used the database for various purposes such as in human risk assessments, worker health and safety assessments, and endangered species, air, and water quality investigations.

The California Department of Pesticide Regulation (DPR) together with the California Department of Food and Agriculture recognize the importance of the PUR data quality. Although the PUR probably provides reasonable estimates of pesticide use for regional and statewide evaluations, there are concerns about data quality for some specific and more localized assessments. There is also a question about the level of compliance with use reporting, i.e., level of under-reporting of pesticide use. In this report we will not attempt to assess the degree of under-reporting. However, we will attempt to understand the quality of the pesticide use data that is collected by DPR.

The purpose of this project was to examine the accuracy of spatial attribute and other data reported in the PUR during the years 1990 to 1997. The error types evaluated for errors or potential errors include duplicate records, inconsistent county codes in the first two digits of the grower identification, meridian/township/range/section (MTRS) values outside county boundaries, missing location identifiers for an agricultural field, inconsistent MTRS values for a geographic location, inconsistent acres planted, and treated acreage greater than acres planted. In addition, we attempted to assess potential errors in the commodity code (also known as site code). However, this last assessment was only partly effective and will be described briefly in this report.

To achieve the objectives of the project, we developed a computer program to check the errors relating to the above data fields. This program used Access 97, ArcView 3.2, and Visual Basic 6.0 software. The error checking application was written mostly in Access Visual Basic Application (VBA), while the GIS maps of the error checking results were produced with ArcView 3.2. Visual Basic provided the front end of the interface between the users and the software. Access and ArcView were used for entering the directory path names for input and output data files. The program allows users to select single or multiple years of pesticide data for error checking.

Error rates were evaluated for PUR data collected from 1990 through 1997. Error rates averaged over this eight-year period were less than 5% for the following error types: duplicate records, inconsistent county code, MTRS outside reporting county, missing location identifiers, inconsistent MTRS for a geographic location, and acres treated greater than the acres planted. In contrast, potential error rates for inconsistent acres planted averaged 8.1% of all agricultural fields and 17% of all agricultural records. In

general, error rates decreased from 1990 to 1997 for inconsistent county codes, MTRS outside reporting county, missing location identifiers, and inconsistent MTRS, while error rates fluctuated from year to year for duplicate records, inconsistent acres planted, and acres treated greater than acres planted. Declining error rates in some of these error types illustrates the effort devoted to improvements in data quality. In addition, low error rates indicate the PUR database is of good quality for many uses.

We also evaluated error-rate distribution spatially by county for this eight-year period. Northern California coastal counties had higher error rates for missing location identifiers than other counties. Mountain areas had higher error rates for inconsistent county codes and MTRS values outside the county boundary. Higher rates for inconsistent acres planted were distributed primarily in the coastal range of southern California and Bay area counties. Lastly, urban and coastal counties generally had higher rates of records with acres treated greater than acres planted. The spatial distribution of these errors may ultimately give us some idea of why they occur.

Although there are errors in the PUR, error rates are generally less than 5% so these data are still useful for many purposes. The results of our analyses provide researchers a sense of how reliable particular data fields are in the PUR. These results should also be useful in prioritizing future efforts to improve the PUR data collection system in both the counties and at DPR.

Introduction

The state of California has required some kind of pesticide use reporting since at least 1950. In 1989, the California Department of Pesticide Regulation (DPR) was given authority to require full pesticide use reporting by the Food Safety Act of 1989 (Chapter 1200, AB 2161) and full use reporting began in 1990. Over two to three and half million records of pesticide use were compiled each year since 1990 (Figure 1; California Department of Pesticide Regulation, 1990-1999). Pesticide use reporting data collected in California is unique nationwide and worldwide (Wauchope and Hornsby, 1992, California Department of Pesticide Regulation, 2000), and is a valuable source of information for many users.

Pesticide use data provide the history of pesticide use for each commodity and pesticide temporally and spatially. The temporal and spatial distribution of the data allows researchers to extrapolate patterns in farming practices, pest population dynamics, human and environmental exposure, as well as loading of pesticides in vulnerable environmental regions. The data have been widely used by staff from federal and state governmental agencies, universities, industrial organizations, environmental groups, and local farmers/citizens for human risk assessments, worker health and safety assessments, and endangered species, air, and water quality studies (Epstein et al., 2000; Domagalski, 1997, 1999; Kratzer, 1997; Troiano and Garretson, 1998; Zhang et al. 1997, 2000; Kegley et al., 2000).

Pesticide use information is compiled into a database, called the Pesticide Use Report, or PUR. Growers and applicators submit use reports to their County Agricultural Commissioner's office. The counties compile the reports and submit these to DPR where the data are error-checked, entered into a database and distributed to interested parties. For their efforts, counties are compensated $0.30 for each record of pesticide use data provided to DPR (California Department of Pesticide Regulation, 2001).

Although regulations and enforcement letters were distributed to each county to ensure that the PUR data were correctly collected and entered into the computer, many types of errors appear in the PUR. Errors enter the PUR in several ways. Many come from inevitable typing mistakes during data entry. Others result from misunderstanding of the meaning of or requirements for the different data fields. These kinds of errors are exacerbated by complicated or confusing requirements and inadequate documentation. Also, different counties may have different data entry programs and different definitions for certain data fields, such as the site location identification.

To use the PUR for any analysis, researchers have used various statistical methods to overcome perceived problems with data quality (Epstein et al., 2000 a, b; Zhang et al., 1997). The diverse statistical methods used for pre-processing PUR data can make it difficult to compare results between various studies. To address the need for a standard data set available to all researchers, DPR and the California Department of Food and Agriculture jointly funded this and other related projects.

The purpose of this project was to examine the accuracy of spatial attribute and other data reported in the PUR during the years 1990 to 1997. The error types studied include:

1. Duplicate records
2. Inconsistent county code in the grower identifier
3. Meridian/township/range/section (MTRS) outside the county boundary
4. Missing location identifiers for an agricultural field
5. Inconsistent MTRS values for a geographic location
6. Inconsistent acres planted for an agricultural field during a growing season
7. Acres treated greater than acres planted
8. Inconsistent site code.

Methods and Criteria

1. Data fields in the PUR database

The PUR contains over 30 database fields. The most important fields (with database field names in parentheses) are pesticide active ingredient (chem._code), pesticide product used (prodno), pounds of active ingredient applied (lbs_chm._used), area or volume treated by this pesticide (acre_treated), area planted (acre_planted), pesticide application method (aer_gnd_ind), grower identification (grower_id), application date (applic_dt), county of application (county_cd), agricultural field location (base_ln_mer, township, tship_dir, range, range_dir, section-these are described below), agricultural field identification (site_loc_id), and crop (site_code).

Each production agricultural record or row in the PUR refers to one active ingredient applied by one application to an individual agricultural field. If a pesticide contains more than one active ingredient, the PUR will contain more than one record. Each non-production agricultural record in the PUR refers to the sum of all applications of an active ingredient by an applicator in a county for one month. In this report, calculations were only made for production agricultural records.

Every record for production agricultural use contains all the information relating to the pesticide application and the geographical location in the Public Land Survey System (PLSS). The PLSS is recorded as meridian/base, township, range, and section (MTRS) and specifies the geographic location of a square-mile area. Thus the PUR only provides agricultural field locations to within a square-mile section.

In the PUR, each grower is identified by a unique string of characters comprising the grower_id, and each agricultural field is also assigned a unique code (site_loc_id) for each grower. However, for this report we assume that an agricultural field is uniquely identified in the PUR by the combination of grower_id, site_loc_id, and MTRS. MTRS is part of the definition of an agricultural field because for the PUR a field must be less than one section (one square mile) in area. An individual agricultural field can receive multiple pesticide applications and therefore the total number of agricultural records is greater than the total number of agricultural fields.

In this study we based our calculations on the number of agricultural fields (distinct values of the combination of grower_id, site_loc_id, and MTRS) and on the number of production agricultural records.

2. Error checking program and its main functions

To achieve the objectives of the project, we developed a computer program to check the errors relating to the above data fields. This program was developed with Access 97, ArcView 3.2, and Visual Basic 6.0 software and requires a Windows 95/98 or Windows NT (4.0 or later) computer with at least 128 MB of RAM and 10GB of hard disk. The error checking application was written mostly in Access VBA, while the GIS maps of the

error checking results were produced with ArcView 3.2. Visual Basic provided the front end of the interface between the users and the software. Access and ArcView were used for entering the directory path names for input and output data files. The program allows users to select single or multiple years of pesticide data for error checking. A detailed description of this program can be found in Appendix I. The program is also available for distribution upon request.

The eight error types evaluated for errors include duplicate records, inconsistent county codes, MTRS values outside the county boundary, missing location identifiers for an agricultural field, inconsistent MTRS values for a geographic location, inconsistent acres planted, acres treated greater than acres planted, and site (commodity) code inconsistency (Diagram 1).

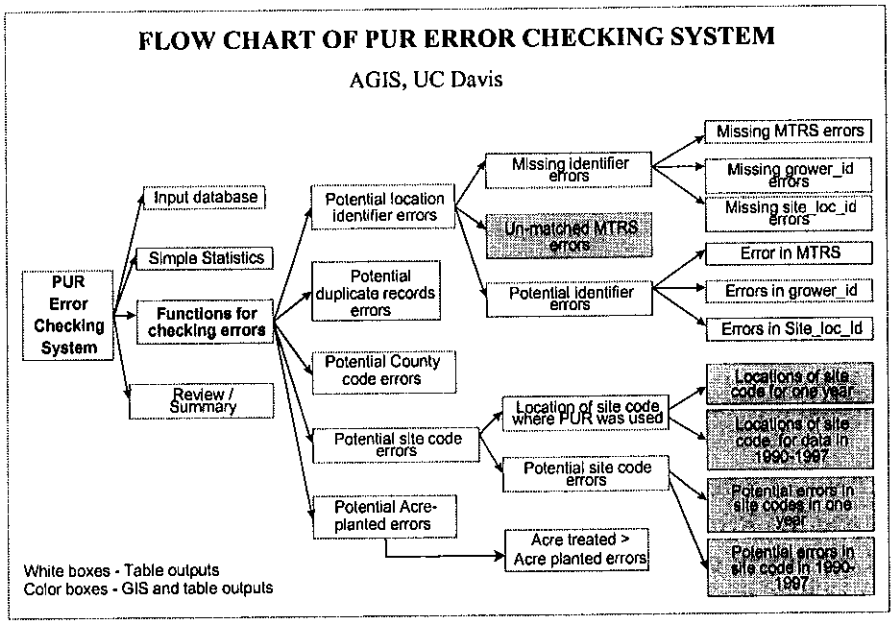


Diagram 1. Chart for the error checking program for spatial attributes.

3. Error types

Duplicate records

Each record in the PUR represents one pesticide application for one chemical active ingredient. Records were considered duplicates of one another if they contained the same values in each of nine data fields: grower_id, site_loc_id, acre_planted, acre_treated, prodno, chem_code, lbs_chm_used, applic_dt, and site_code (See Appendix I for details).

Duplicate records were extracted from the PUR database and saved in a separate file for later examination (see Appendix I for file conventions). The number of duplicate records in the saved file is used to estimate the potential duplicate error rate for each county and each year. If two or more records were found with identical values for the nine data fields, only one record was counted when calculating the duplicate error rate.

Inconsistent county code

In the grower identification string, the first two digits in the grower_id field signify the reporting county code. This should be the same as the county code reported for the record. If the first two digits were not the same as the county code, these records were saved into a separate file for further examination. For example, Yolo County has a code of 57. The program will find any records submitted from Yolo County that do not have "57" in the first two digits of the grower_id field, and will save these records in a separate file. The number of records in the saved file was used to estimate the error rate for inconsistent county code.

MTRS outside the county boundary

The MTRS value reported for an agricultural field in the PUR should be within the reporting county's boundary. MTRS values outside the reporting county were treated as errors. The program compares the MTRS from the PUR to a PLSS GIS county map. Records with MTRS errors were then saved into a separate file for later use. The number of distinct MTRS values that occurred in a record where the MTRS was outside of the reporting county's boundary was used to estimate the error rate. The error rate was calculated in two different ways: 1) 100 times the number of distinct MTRS values outside of the county boundary divided by the number of all agricultural fields and 2) the percent of all agricultural production records with MTRS values outside the county.

Missing location identifiers for an agricultural field

An agricultural field is defined here as a unique combination of three data fields: grower_id, site_loc_id, and MTRS. A missing location identifier refers to missing information in the PUR for any of one of these data fields. If one or more location identifiers were missing in that record, the record was saved in a different file for future use. The number of records with missing location identifiers was used to estimate the error rate.

Inconsistent MTRS for a geographic location

To perform this calculation, we assumed that the combination of grower_id and site_loc_id uniquely identified a single geographic location within a section. Under this assumption, all records with applications to this location should have the same MTRS value. Records with different MTRS values for the same combination of grower_id and site_loc_id are considered inconsistent and are considered possible errors. These records were saved in a file and the number of saved records was used in estimating the error rate. The error rate was calculated in two different ways: 1) 100 times the number of distinct grower_id and site_loc_id values with inconsistent MTRS values divided by the number of all agricultural fields and 2) the percent of all production agricultural records with inconsistent MTRS values.

Inconsistent acres planted

The value of acres planted should be the same for an agricultural field during one growing cycle for one crop. However, acres planted reported in the PUR were sometimes different at different times for one field and one crop. Again, each agricultural field is identified by the grower_id, site_loc_id, and MTRS, but we considered acres planted to be inconsistent for a field only if the acres planted differed between records for the same agricultural field and the same crop. We saved those records with inconsistent acres planted values into a separate file. We calculated the error rate in two different ways: 1) 100 times the number of agricultural field and crop combinations which had inconsistent acres planted divided by the number of agricultural fields and 2) the percent of all production agricultural records with inconsistent acres planted.

Acres treated greater than acres planted

Within each PUR record, the reported acres treated should be smaller than or equal to the acres planted. If the acres treated are greater than the acres planted in any record, then either the acres treated or the acres planted is wrong. The error checking program extracted the records with acres treated greater than the acres planted after the inconsistent acre planted records were removed. These records were then saved to a file and later used to estimate the error rate as a percent of the total number of production agricultural records.

Site code

Site code in the PUR identifies the commodity that receives a pesticide application. Together with grower_id, site_loc_id and MTRS, site code can help identify the location of pesticide applications reported in the PUR. Site code is another important data field to be considered in spatial aspects for the PUR. To get some idea of the data quality for site code, we compared the PUR commodity locations with the crop maps from the California Department of Water Resources (DWR). DWR had maps for the major crops in 16 central valley counties. But these DWR maps were only produced for certain years, i.e., only one year for one or two counties. We made comparisons for the following 10 commodities: almonds, grapes, cotton, rice, alfalfa, tomatoes, walnuts, oranges, lettuce, and broccoli.

The comparison between site code from the PUR and crops from DWR maps was made in the interface of ArcView and Access. We prepared two tables, each containing site codes and all MTRS values where those crops were reportedly grown. One table contained the values from the PUR and the other values from the DWR crop maps. The tables contained data only for those crops that appeared in the DWR crop maps. Then we compared the two tables. If any discrepancies occurred, the program generated a difference map and displayed it on the screen in ArcView. Users can save the maps and/or data files. Differences could result from errors in either the PUR or DWR and either in the site code or location of the crop. Since land-use maps exist for only one year for any county, we could not analyze error rate trends. However, information concerning

potential errors for the commodities mentioned above is available upon request.

Results

Duplicate records

The proportion of duplicate records averaged 2.3% of all agricultural records for the eight-year period (Table 1). Annually, the percent of duplicate records ranged from 1.4 to 4.1% from 1990 to 1997 (Table 1, Figure 2). The highest duplicate record rate, 4.1%, occurred in 1993, and the lowest rate, 1.4%, in 1996. There appears to be a slight decreasing trend in percent of duplicate records after 1993.

The percent of duplicate records were highest in Imperial, Monterey, Santa Cruz, Sonoma, and Mono Counties (Figure 3). However, in each of these counties, the percent of duplicate records was unusually high only during one year. In all other years, the error rates were similar to the rates in the other counties, mostly less than 3%.

There were no clear trends in percent of duplicate records within the different counties, with percentages fluctuating from year to year (Table 2). However, the high duplicate record rate in 1993 was dominated by Monterey County, which was an unusually high 16%.

Inconsistent county code

The proportion of records with inconsistent county codes averaged <1% of all agricultural records for the eight-year period (Table 1). Annually, error rates ranged from near 0 to 4.5% from 1990 to 1997 (Table 1). The error rate for inconsistent county code decreased dramatically from 1990 to 1992 (Table 1, Figure 4). The inconsistent county code error rate was less than 0.04% for each year from 1993 to 1997 although the highest average error rate was 4.5% in 1990.

Most counties had very low error rates except for San Benito County and those in the mountain regions where agriculture is less important (Figure 5). During 1995 to 1997 most counties had no errors in county code (Table 3). The only exceptions were 1 error in Sutter in 1995, 1 error in Santa Barbara in 1996, and 74 errors in Napa in 1997.

MTRS outside the county boundary

The error rate for MTRS values outside the county boundary averaged <0.06% of all agricultural fields for the eight-year period (Table 1). Annually, error rates decreased from 0.4% in 1990 to nearly 0 in 1997 (Table 1).

The error rate for MTRS values outside the county boundary averaged 0.79% of all agricultural records for the eight-year period (Table 1). Annually, error rates decreased from 4.3% in 1990 to 0.001% in 1997 (Table 1, Figure 6). Most of the errors appeared in the mountain area and some coastal counties (Table 4, Figure 7). Although some of the mountain area counties had high error rates, above 10%, agricultural activities in these counties were minimal. The error rates for most of the valley counties were lower than 3% during 1990 to 1997.

Missing location identifiers

Location identifiers include the data fields: grower_id, site_loc_id and MTRS. The error rate for missing location identifiers averaged 0.004% of all agricultural production records for the eight-year period. Annually, error rates decreased from 0.01% in 1990 to 0.0006% in 1997 (Table 1, Figure 8). Most of the missing location identifiers were in the northern coastal counties (Figure 9). The highest average error rates, 0.04 to 0.17%, were in the counties of Mono, Mendocino and Marin (Figure 9). However, these counties had few agricultural records and the total number of records in these counties with errors was rather small during 1990 to 1997. In 1994, nearly all of these errors were in Mendocino (Table 5). Most of the valley counties had error rates less than 0.01%.

Inconsistent MTRS for a geographic location

The error rate for inconsistent MTRS values averaged 1.66% of all agricultural fields over the eight-year period. Annually, error rates calculated as a proportion of total number of agricultural fields ranged from a high of 3.1% in 1990, to a low of 0.4% in 1995 (Table 1, Figure 10).

The error rate for inconsistent MTRS values averaged 4.6% of all agricultural production records for the eight-year period. Annually, error rates for inconsistent MTRS calculated as a proportion of total number of agricultural production records ranged from a high of 8.6% in 1990, to 1.1% in 1995 (Table 1, Figure 12). In general, error rates decreased during the eight-year period.

San Luis Obispo, Ventura, and Los Angeles counties had the highest error rates (both by agricultural fields and PUR records) during 1990 to 1997, with rates above 10% for most years. However, in all three counties, the error rates generally decreased from 1994 to 1997 (Tables 6 and 7, Figures 11 and 13). Most of the counties in the state had error rates less than 5%.

Inconsistent acres planted

Potential error rates for inconsistent acres planted averaged 8.1% of all agricultural fields over the eight-year period. Annually, potential error rates ranged from a high of 20% in 1992, to a low of 11% in 1990 (Table 1, Figure 14). Potential error rates for inconsistent acres planted averaged 17% of all agricultural production records over the eight-year period. Annually, potential error rates for inconsistent acres planted calculated as a

proportion of total number of agricultural production records ranged from a high of 10% in 1992, to a low of 11% in 1990 (Table 1, Figure 16). Potential error rates for this error type varied from year to year with no increasing or decreasing trend.

Higher error rates appeared in southern coastal and bay areas counties, especially in Ventura, San Luis Obispo, Santa Barbara, Monterey, and San Benito Counties, where error rates exceeded 30% in many years (Tables 8 and 9, Figures 15 and 17). Although error rates remained nearly constant for most counties from 1993 to 1997, error rates decreased in many valley counties (Tables 8 and 9).

Acres treated greater than acres planted

The proportion of records with acres treated greater than acres planted averaged 0.72% of all agricultural records for the eight-year period (Table 1). Annually, error rates ranged from a high of 1.4% in 1991, to a low of 0.30% in 1995 (Table 1, Figure 18). The error rate varied from year to year with no increasing or decreasing trend.

The highest error rates were in the Bay Area counties and Los Angeles County (Figure 19). However, the highest errors in these counties occurred in one or two years during 1990 to 1997 (Table 10). These counties include Alameda, Contra Costa, Los Angeles, Marin, Orange, San Mateo, and Santa Clara counties.

Site code

The discrepancies between the site code in the PUR and the crop/land use data from DWR were saved for 16 central valley counties that have available crop/land use digital files. Results of this analysis are available for distribution upon request.

Discussion

Statewide, error rates averaged for the eight-year period (1990-1997) were less than 5% for the following error types: duplicate records, inconsistent county code, MTRS outside reporting county, missing location identifiers, inconsistent MTRS for a geographical location, and acres treated greater than the acres planted. In contrast, the statewide average error rates for inconsistent acres planted were 8.1% of all agricultural fields and 17% of all agricultural records. In general, error rates decreased for inconsistent county codes, MTRS outside reporting county, missing location identifiers, and inconsistent MTRS, while error rates fluctuated from year to year for duplicate records, inconsistent acres planted, and acres treated greater than acres planted.

The decline in error rate of some error types demonstrates the effort that has been devoted to improvements in data quality. In addition, low error rates indicate the PUR database is of good quality for many uses.

Error rates for inconsistent county codes, MTRS outside of county boundaries, and missing location identifiers in the PUR declined almost to 0% in 1997. It is possible this decline is a result of increased experience and familiarity of growers, pesticide applicators, and county agricultural staff with county codes, MTRS designations, and field location identifiers. Although average error rates for most of these error types are rather small, it is still important to consider them when analyzing pesticide impacts or trends in a localized area.

Most of these errors could be caught in the data entry programs used by the counties. Errors indicate areas where the data entry programs can be improved, such as identifying incorrect MTRS values in certain counties. Incorrect MTRS values in the PUR show that either the county data entry program contained the wrong MTRS information or the county staff altered the values for MTRS during data entry.

Not all of the records found for these different error types are actually errors, especially for duplicate records and inconsistent acres planted. For example, if a grower separately reported individual spot treatments to an agricultural field with the same active ingredient, pounds applied, and area treated, the records in the PUR would appear to be duplicates when in fact they are not. Meanwhile, erroneous duplicates can occur in several ways. For example, when two people are responsible for pesticide applications on a farm and both send in a report for the same application, a duplicate record will result. Or, when the same report unknowingly gets entered into the database two or more times, a duplicate record will result. Therefore, the error rate calculated in this study is only an estimate of the actual error rate.

Duplicate records directly influence the estimate of the total amount of pesticide used in a county, region, or state. Many studies reference the PUR as the pesticide use data source when assessing pesticide impacts on the environment (Troiano, 1998; Domagalski, 1997). Reducing the number of erroneous duplicate records is an essential step in providing accurate impact assessments. To reduce duplicate record errors, better guidelines should be developed such as always reporting actual acreage treated for spot treatments.

Inconsistency in acres planted is one of the data fields in the PUR with a potentially high error rate. Not all of these are necessarily errors and it may be that in some situations most of them are not errors. In this report we distinguished agricultural fields by grower_id, site_loc_id, and MTRS. The potential errors for acres planted in this report were considered crops growing on the field. If the agricultural field grows multiple crops in a year, the number of fields in this case will be more than the actual physical number of agricultural fields. Therefore, the high potential error rate in acres planted may reflect multiple cropping as well as actual errors. It is important that we pay attention to the trends of these error rates rather than the absolute values.

In the real world, an agricultural field is an individual physical contiguous area of land that undergoes the same cultural practices. Ideally, we would hope that each grower

assigns a unique and distinct site_loc_id to each separate agricultural field. MTRS is part of our definition of an agricultural field because for the PUR a field must be less than one section (one square mile) in area and sometimes a grower will assign one site_loc_id to a field larger than one square mile. Also, some growers apparently assign only one site_loc_id to all their fields. In this case the only way to know from the PUR that they are different fields is if they are in different sections and this requires the use of the MTRS value to distinguish the fields.

However, including MTRS in the definition of an agricultural will overestimate the number of fields if errors occur in assigning MTRS. The magnitude of this problem, however, can be seen from the error rates for inconsistent MTRS. That error rate was found by looking for inconsistent MTRS values for each distinct combination of grower_id and site_loc_id. That error rate was much lower than the error rate for inconsistent acres planted so including MTRS in the definition of agricultural field seems reasonable.

However, if a grower uses the same site_loc_id for several distinct fields all in one section and if two or more fields are different sizes but with the same crop, then these would appear as inconsistent acres planted. However, these inconsistent acres planted values would not actually be errors.

In other situations, these inconsistent acres planted are errors. At the beginning of each calendar year, growers who plan to apply pesticides must get a permit or an operator identification from their county. When applying for this form, the grower indicates the anticipated crops and acreage planned for the coming year. It is this estimated acreage that is first entered into the pesticide use report database. There are a few potential sources of errors for this data field. One error could occur when the person requesting the form enters the wrong acreage. Another error could occur during data entry. A third type of error could occur if changes are made to the amount of acreage planted for a particular crop. For various reasons either economic or climatic considerations, the grower may alter the acreage actually planted during the season. Therefore, it is possible that the value of acres planted originally reported may differ from the actual acres of an agricultural field at some point in the season. When changes in acres planted occur during the year, the acres planted originally reported to the county is not necessarily corrected retroactively. Therefore, without further information from each county, we can only consider these potential errors in acres planted.

The method we used to calculate percent of errors based on the number of agricultural fields artificially increases the values. We calculated the number of distinct values of grower_id, site_loc_id, MTRS, and site_code which had inconsistent acres planted and divided by the number of distinct values of grower_id, site_loc_id, and MTRS (our definition of an agricultural field). If many fields had several different crops planted on them during a season, then it would even be possible for this ratio to be greater than 1. For this error rate, it would be better to divide by the total number of distinct values of grower_id, site_loc_id, MTRS, and site_code. However, to avoid the confusion of the definition of an agricultural field, we used the same agricultural field values for this

calculation. The error rates for counties where a high proportion of agricultural fields grow multiple crops in one year are expected to be higher than for counties that grow a high proportion of annual and perennial crops. Since the PUR does not contain multiple cropping data, we cannot further separate the information.

If the fields with inconsistent acres planted happened to have a large number of applications, then the record error rate would be increased because all records for these applications would be marked as inconsistent. This might explain why the record error rate was consistently higher than the agricultural field based error rate.

Knowing the rates of various kinds of errors is important in helping PUR users determine what kinds of analyses are more or less reliable. It is clear that the highest error rates are associated with acres planted, which is important in many kinds of analyses. Implementing the Food Quality Protection Act requiring estimates of percent of the planted acres of a crop that are treated is one example. Accurate estimates depend on accurate data on acres treated and acres planted.

Thus, knowing these potential error rates is also important in helping DPR determine where the PUR collection system needs improvement. Further study is required to more clearly understand which of these potential errors are really problems, to develop methods to correct or minimize the errors in analyses, and to improve the PUR collection process. The error rates we have reported here will help us choose the kinds of errors and particular counties with the largest problems for further investigation. In phase 2 of the overall error-checking project, we will report on the results of a survey sent to each county to document their PUR collection procedures and definitions. We will use the historical error rates to help us determine which data fields need further study through the survey and its follow up.

Although this report documents the potential error rates for various data fields in the PUR, the PUR is one of the best databases in the world for tracking pesticide use. Most of the errors could easily be caught and fixed at the county during data entry. This database has been and will continue to be widely used for risk assessment, pest management, protecting air and water resources and marketing research. Therefore we should strive to improve data quality wherever possible.

Acknowledgement

The authors wish to acknowledge Dr. Lisa Ross for the useful comments in reviewing this report. We want to thank the visiting scholars at the AGIS laboratory at the University of California Davis for their time and hard work in producing results of this study. We are thankful to Dr. Charlie Goodman and Dr. David Supkoff for earlier discussion on this project. We want to thank the California Department of Food and Agriculture for the financial support for this study.

References

California Department of Pesticide Regulation. 2001. Pesticide use reporting County Contract. Sacramento, CA.

California Department of Pesticide Regulation. 2000. Pesticide use reporting: an overview of California's unique full reporting system. Sacramento, CA.

California Department of Pesticide Regulation. 1990-1999. Summary of pesticide use report data indexed by Chemical and by commodity. California Department of Pesticide Regulation, Sacramento, CA.

Domagalski, J. 1997. Pesticides in surface and ground water of the San Joaquin – Tulare Basins, California: Analysis of available data, 1966 through 1992. US Geological Survey, Water-Supply Paper 2468.

Epstein, L. Bassein, S., Zalom, F. G. and L.R. Wilhoit. 2000 a. Reduction in use of organophosphates in almond orchards during the rainy season in California. Agriculture, Ecosystem and Environment. (in press).

Epstein, L., Bassein, S. and F. G. Zalom. 2000 b. Almond and stone fruit growers reduce OP, increase pyrethroid use in dormant sprays. California Agriculture. 54:14-19.

http://www.cdpr.ca.gov/dprdatabase.htm. PUR database descriptions and related information. California Department of Pesticide Regulation, Sacramento, CA.

Kegley, S.E., Orme, S. and L.N. Neumeister. 2000. Hooked on poison: pesticide use in California 1991-1998. Pesticide Action Network and California for Pesticide Reform, San Francisco, CA. p84.

Kratzer, C. R. 1997. Transport of Diazinon in the San Joaquin River Basin, California. US Geological Survey, Open-file Report 97-411.

Kuivila, K.M. and C. G. Foe. 1995. Concentrations, transport and biological effects of dormant spray pesticides in the San Francisco estuary, California. Environmental Toxicology and Chemistry. 14: 1141-1150.

Troiano, J. and C. Garretson. 1998. Movement of Simazine in runoff water from citrus orchard row middles as affected by mechanical incorporation. Journal of Environmental Quality. 27: 488-494.

Wauchope, D. and A. Hornsby. 1992. The SCS/ARS/CES pesticide properties database for environmental decision-making. Reviews of Environmental Contamination and Toxicology 123: 1-164.

Zhang, M., Geng, S., Ustin, S. and K. K. Tanji. 1997. Pesticide occurrence in groundwater in Tulare County, California. Environmental Monitoring and Assessment 45: 101-127.

Zhang, M., Smallwood, Shawn and Erin Anderson. 2001. Relating indicators of ecological health and integrity to assess risks to sustainable agriculture and native biota. In Managing Ecosystem Health. (eds. Alexander H. Harcourt), Vol 2, "Issues and Methods", Section 6 "Communities, Politics, Culture, and Tradition". (in press)

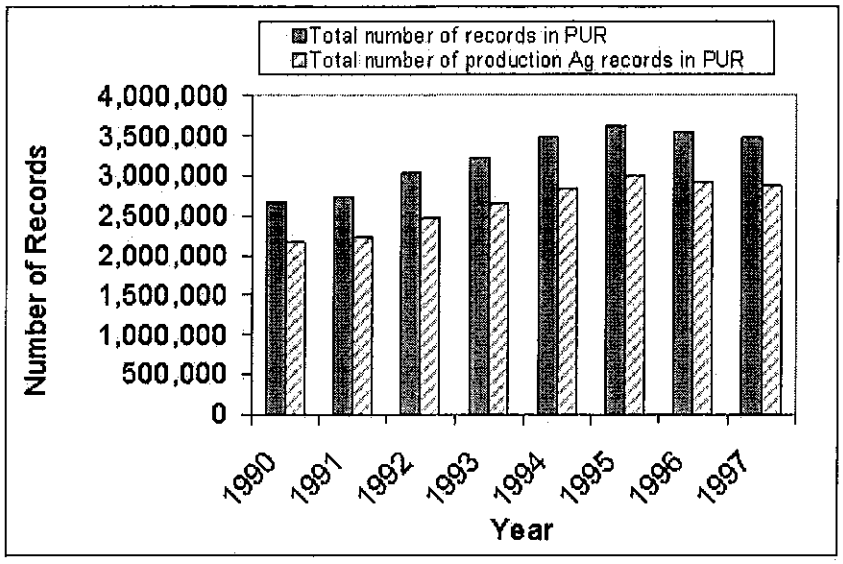


Figure 1. Total number of records in PUR (1990 to 1997)

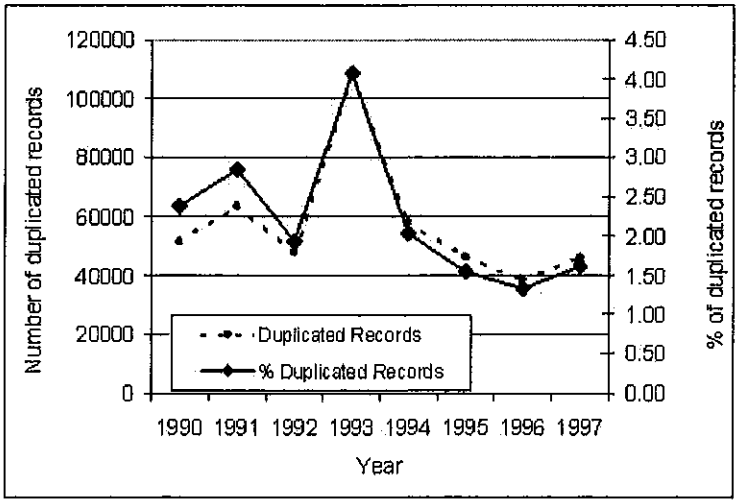


Figure 2. Number and percent of production Ag duplicate records in PUR 1990-1997

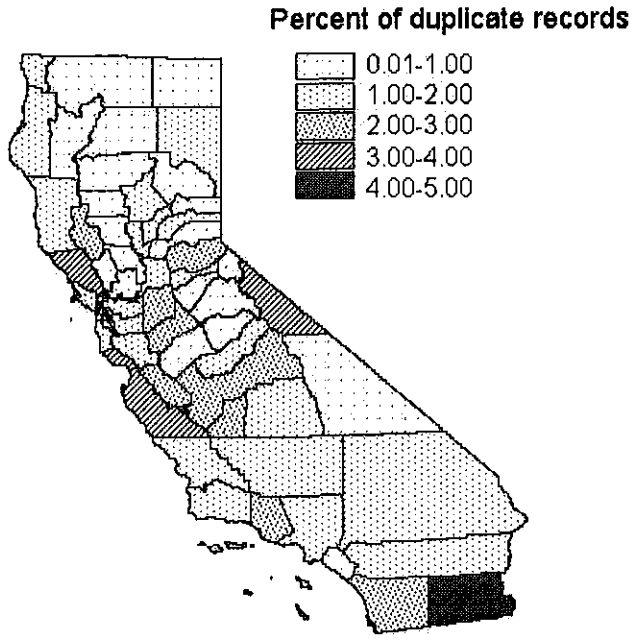


Figure 3. Spatial distribution of the county average percent of duplicate records in PUR 1990-1997.

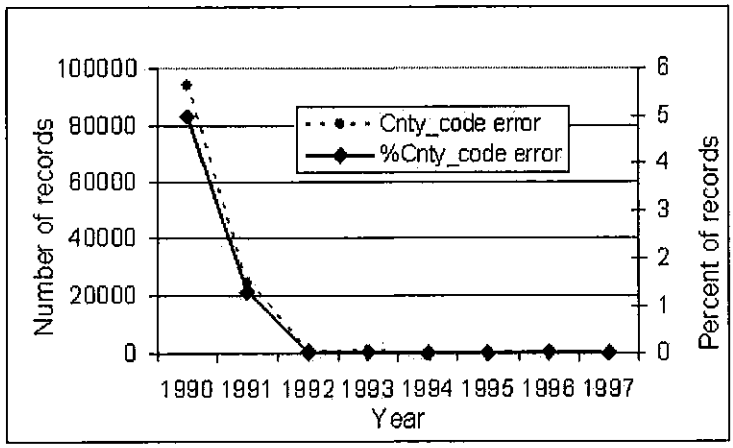


Figure 4. Number and percent of production agricultural records with inconsistent county codes in grower_id.

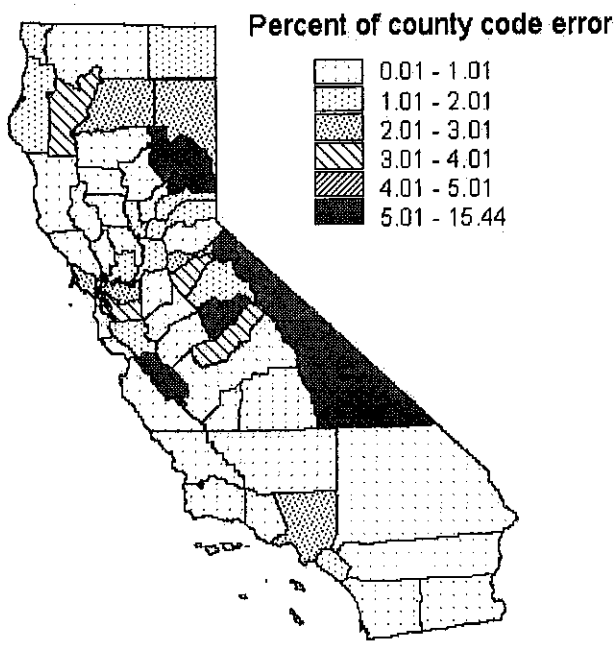


Figure 5. Spatial distribution of the county average percent of county code errors in grower_id 1990-1997.

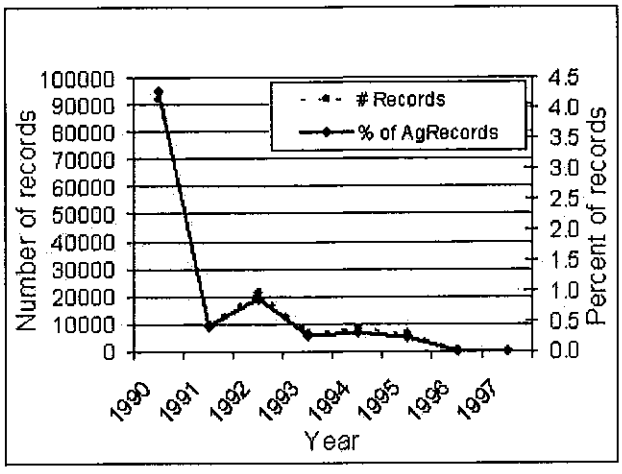


Figure 6. Number and percent of production agricultural records with of township/range/section (MTRS) outside county boundaries.

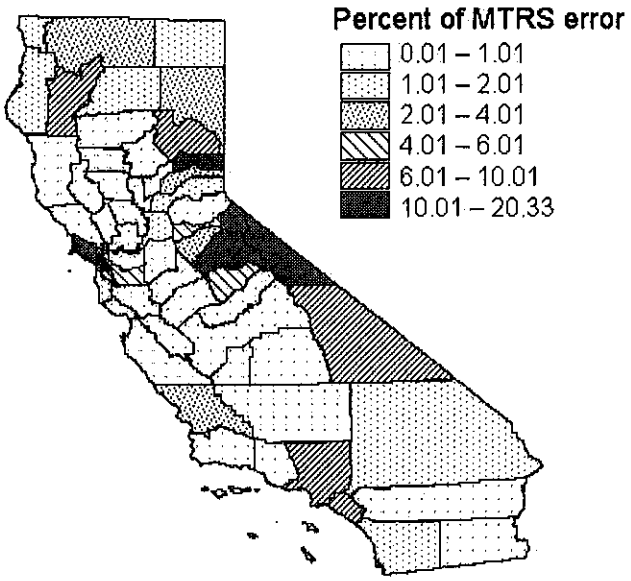


Figure 7. Spatial distribution of the county average percent of township /range/section (MTRS) records outside county boundaries (1990 -1997).

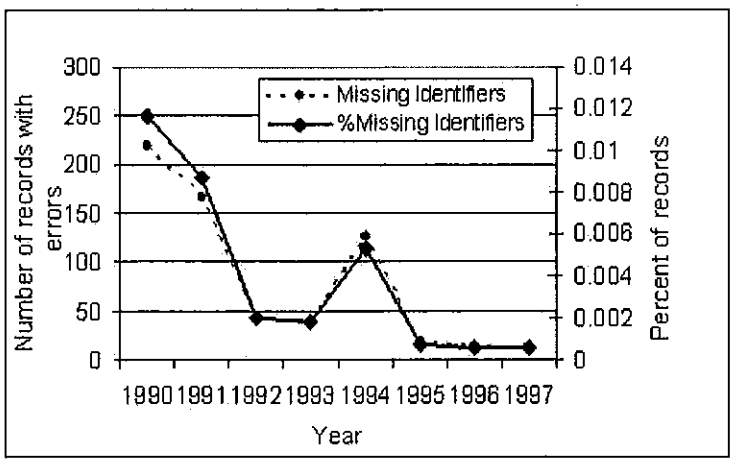


Figure 8. Number and percent of production agricultural with missing location identifiers (MTRS, grower_id, and site_loc_id) 1990-1997.

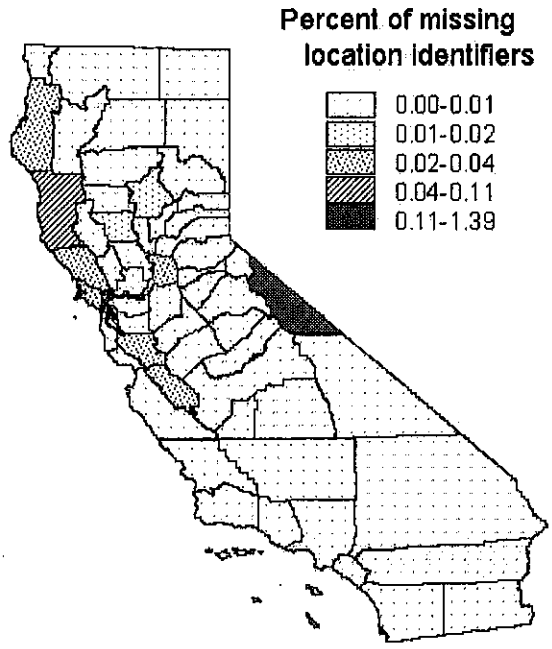


Figure 9. Spatial distribution of the county average percent of missing location identifiers (MTRS, grower_id, site_loc_id) 1990-1997.

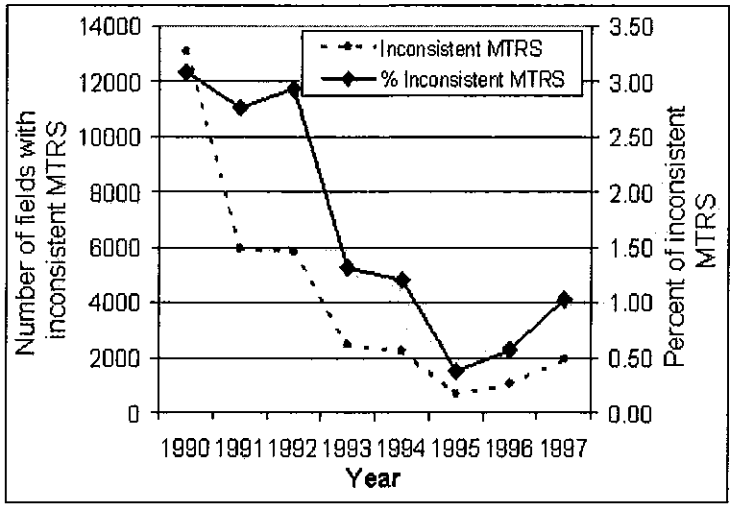


Figure 10. Number and percent of agricultural fields with inconsistent MTRS for grower_id and site_loc_id 1990-1997

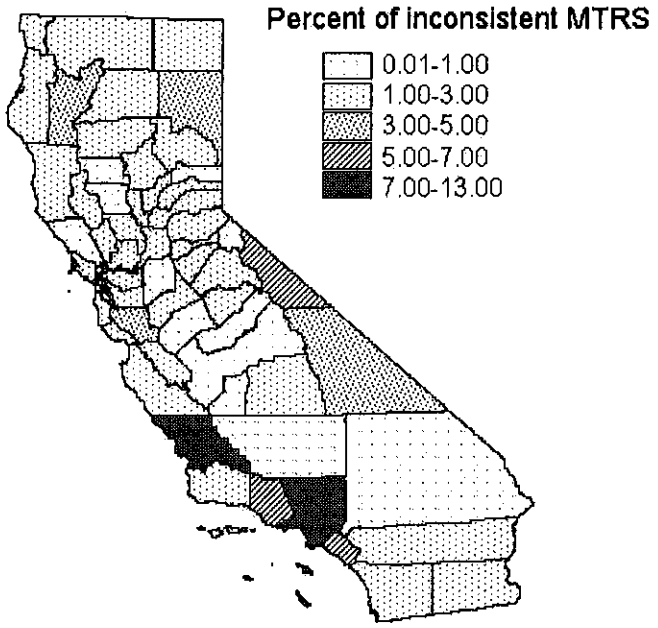


Figure 11. Spatial distribution of the county average percent of fields with inconsistent MTRS for grower_id and site_loc_id 1990-1997

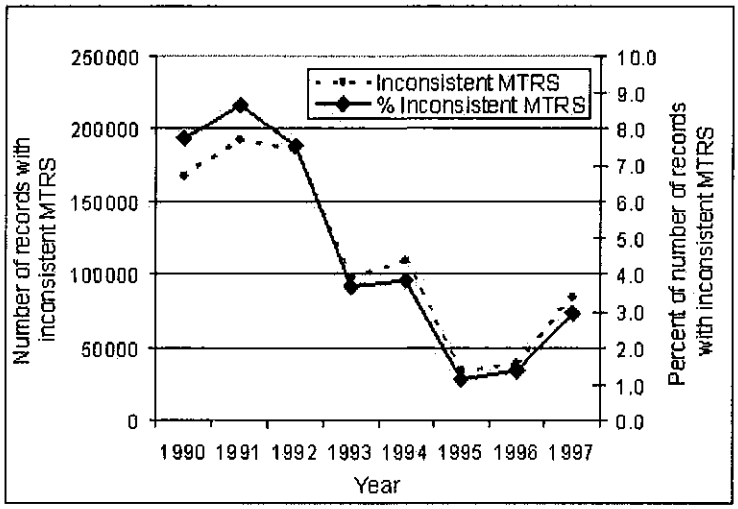


Figure 12. Number and percent of production agricultural records with inconsistent MTRS for grower_id and site_loc_id

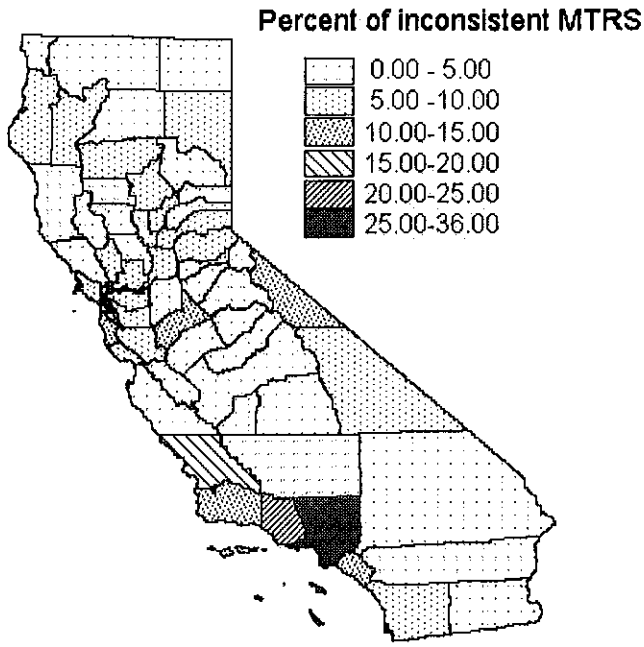


Figure 13. Spatial distribution of the county average percent of records with inconsistent MTRS for grower_id and site_loc_id 1990-1997

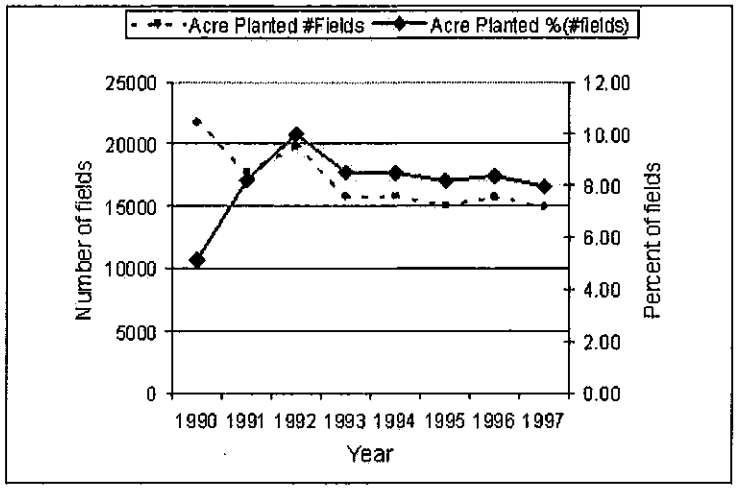


Figure 14. Number and percent of agricultural fields inconsistent acres planted

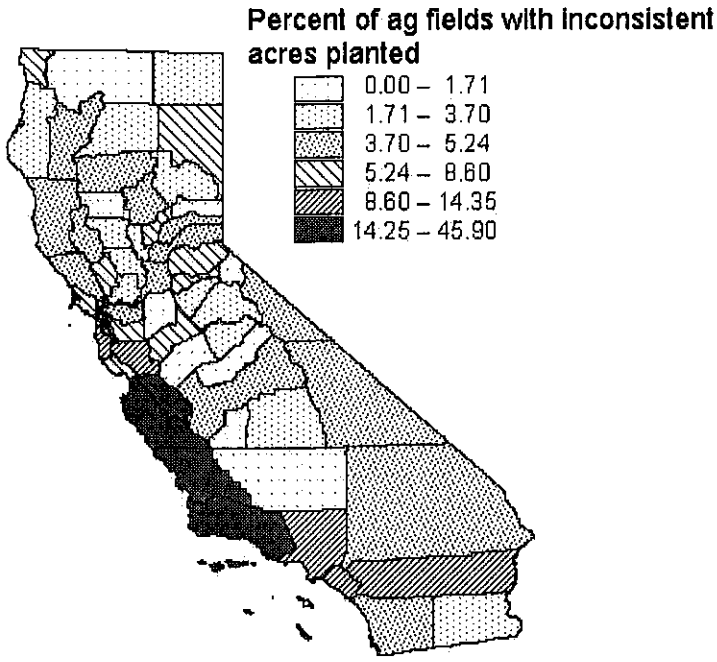


Figure 15. Spatial distribution of the county average percent for agricultural fields with inconsistent acres planted 1990-1997.

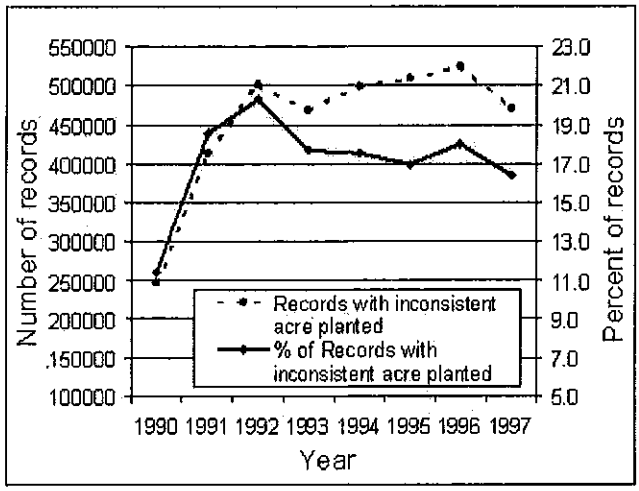


Figure 16. Number and percent of production agricultural records with inconsistent acres planted.

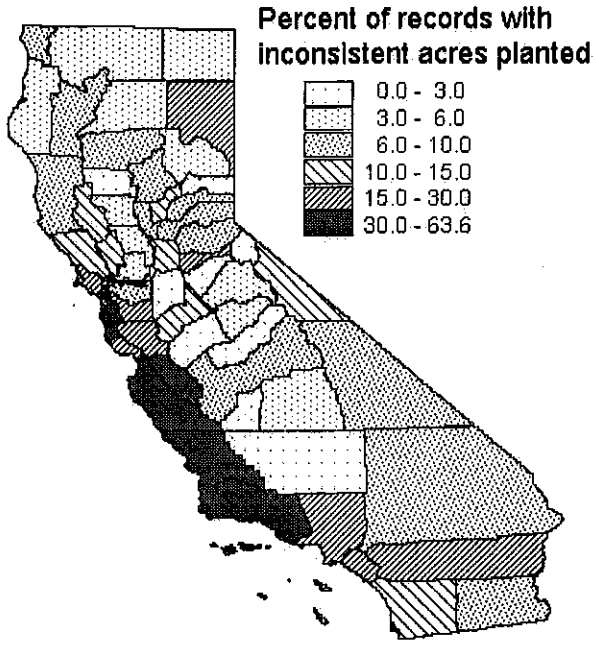


Figure 17. Spatial distribution of the county average percent of the records with inconsistent acres planted 1990-1997.

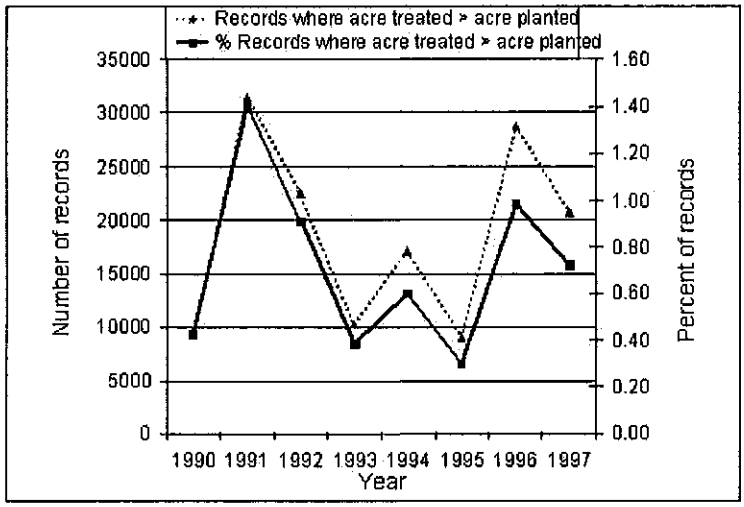


Figure 18. Number and percent of production agricultural records where acre treated was greater than acre planted.

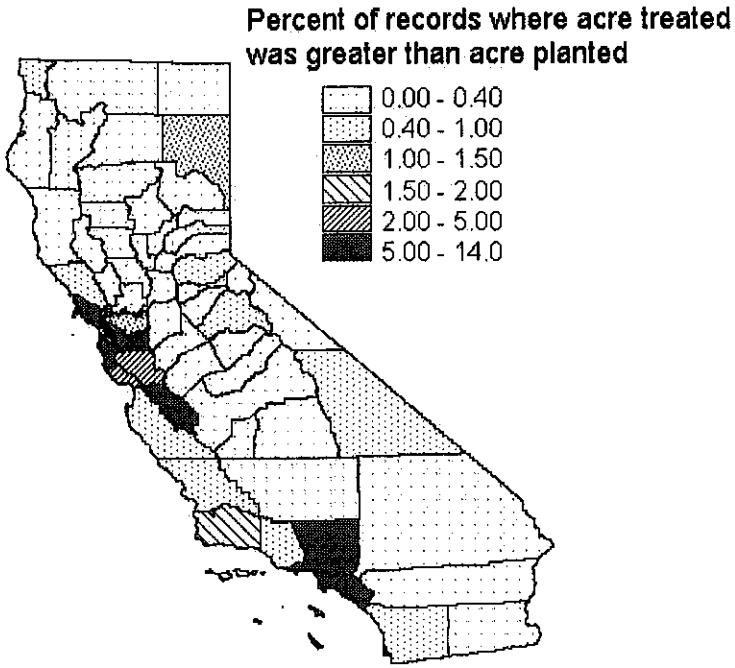


Figure 19. Spatial distribution of the county average percent of records where acre treated was greater than acre planted 1990-1997.

Table 1. Number and percent of errors for selected error types in the PUR from 1990 to 1997.

	1990	1991	1992	1993	1994	1995	1996	1997	Average
Duplicate Records									
# Records	52376	64963	48670	109074	60337	48629	39527	46235	58726
% of AgRecords	2.423	2.912	1.972	4.122	2.119	1.618	1.355	1.611	2.267
Inconsistent County Code									
# Records	96327	26261	676	926	9	1	468	74	15593
% AgRecords	4.457	1.177	0.027	0.035	0.000	0.000	0.016	0.003	0.714
MTRS Outside the County Boundary									
# MTRS	8751	585	534	254	189	129	30	4	1310
# Records	92310	9165	21527	6732	8630	6561	194	27	18143
% AgFields	0.405	0.026	0.022	0.010	0.007	0.004	0.001	0.000	0.059
% AgRecords	4.271	0.411	0.872	0.254	0.303	0.218	0.007	0.001	0.792
Missing Location Identifier for an Agricultural Field									
# Records	219	167	42	40	127	19	15	14	80
% AgRecords	0.0116	0.0087	0.002	0.0018	0.0053	0.0007	0.0006	0.0006	0.004
Inconsistent MTRS for a Geographic Location									
# Fields	13068	5979	5850	2461	2250	702	1082	1947	4167
# Records	167293	192962	185951	97827	109797	34197	39850	84869	114093
% AgFields	3.085	2.765	2.938	1.320	1.207	0.380	0.573	1.035	1.663
% AgRecords	7.741	8.650	7.536	3.697	3.856	1.138	1.366	2.957	4.617
Inconsistent Acres Planted									
# Fields	21790	17773	19863	15849	15800	15095	15769	14989	17116
# Records	247014	413931	501255	468807	498653	509467	525075	470385	454323
% AgFields	5.144	8.218	9.975	8.499	8.478	8.165	8.353	7.970	8.100
% Ag Records	11.429	18.555	20.314	17.718	17.514	16.950	17.994	16.388	17.108
Acres Treated Greater Than Acres Planted									
# Records	9260	31367	22464	10216	17092	8993	28735	20694	18603
% AgRecords	0.428	1.406	0.910	0.386	0.600	0.299	0.985	0.721	0.717

Table 2. The number and percent of duplicate records reported in the PUR, summarized by county and year.

Cnty_Name	Data	1990	1991	1992	1993	1994	1995	1996	1997
Alameda	# of Records	120	30	0	15	186	269	14	259
	% of Ag Records	1.33	0.41	0.00	0.28	2.16	3.81	0.29	3.29
Alpine	# of Records	6	0	0	0	0	0	0	0
	% of Ag Records	1.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Amador	# of Records	36	134	9	16	17	52	0	0
	% of Ag Records	1.74	4.33	0.53	0.83	1.09	2.49	0.00	0.00
Butte	# of Records	41	377	560	708	610	373	427	396
	% of Ag Records	0.18	1.43	1.97	2.42	1.62	1.07	1.25	1.23
Calaveras	# of Records	11	4	6	7	1	0	0	0
	% of Ag Records	1.26	0.68	0.84	0.78	0.10	0.00	0.00	0.00
Colusa	# of Records	249	187	207	301	81	194	44	36
	% of Ag Records	1.21	1.18	0.96	1.28	0.34	0.72	0.18	0.17
Contra Costa	# of Records	97	44	73	123	110	140	98	103
	% of Ag Records	1.51	0.59	0.88	1.21	1.17	1.46	0.95	0.80
Del Norte	# of Records	288	8	6	70	31	28	71	75
	% of Ag Records	7.73	0.27	0.23	1.76	0.87	0.76	1.81	1.76
El Dorado	# of Records	150	47	92	38	112	1	0	0
	% of Ag Records	5.80	2.04	3.44	1.32	3.88	0.05	0.00	0.00
Fresno	# of Records	7024	19723	5340	6346	8274	9810	7116	9437
	% of Ag Records	2.54	7.37	1.67	1.85	2.31	2.31	2.00	2.50
Glenn	# of Records	167	702	205	57	47	49	33	99
	% of Ag Records	0.84	3.35	1.03	0.28	0.20	0.21	0.15	0.42
Humboldt	# of Records	50	65	27	99	49	25	12	27
	% of Ag Records	1.27	1.78	0.75	3.83	1.88	0.90	0.41	1.10
Imperial	# of Records	2193	15236	1047	2478	1636	2196	2025	2138
	% of Ag Records	2.89	15.91	1.85	3.31	1.94	2.28	2.14	2.32
Inyo	# of Records	0	0	1	0	0	0	0	0
	% of Ag Records	0.00	0.00	1.18	0.00	0.00	0.00	0.00	0.00
Kern	# of Records	19157	632	842	552	1178	1208	935	1377
	% of Ag Records	11.88	0.50	0.55	0.34	0.68	0.58	0.54	0.80
Kings	# of Records	463	2218	564	1662	1795	556	542	4507
	% of Ag Records	1.03	4.31	1.00	2.67	2.47	0.92	0.74	5.61
Lake	# of Records	283	993	321	161	8	5	0	0
	% of Ag Records	4.63	10.82	4.25	1.90	0.21	0.06	0.00	0.00
Lassen	# of Records	4	0	2	2	6	2	2	27
	% of Ag Records	1.26	0.00	0.42	0.59	1.37	0.57	0.42	4.40
Los Angeles	# of Records	212	117	65	147	248	116	249	204
	% of Ag Records	1.97	1.23	0.59	1.22	1.81	1.11	1.64	1.41
Madera	# of Records	880	1085	2375	1171	241	183	60	283
	% of Ag Records	1.48	1.80	3.50	1.64	0.35	0.25	0.09	0.43
Marin	# of Records	12	69	1	19	3	21	4	0
	% of Ag Records	1.66	6.06	0.08	1.65	0.22	1.81	0.90	0.00
Mariposa	# of Records	3	0	1	7	0	0	0	0
	% of Ag Records	1.85	0.00	0.61	2.26	0.00	0.00	0.00	0.00
Mendocino	# of Records	200	118	157	125	27	48	8	51

	% of Ag Records	2.54	1.43	1.78	1.29	0.27	0.44	0.08	0.48
Merced	# of Records	806	80	41	49	2	0	3	22
	% of Ag Records	1.04	0.10	0.04	0.04	0.00	0.00	0.00	0.02
Modoc	# of Records	7	28	117	57	2	2	3	10
	% of Ag Records	0.35	1.30	3.49	1.63	0.08	0.05	0.08	0.22
Mono	# of Records	11	1	0	0	4	4	0	0
	% of Ag Records	4.95	1.75	0.00	0.00	4.00	15.38	0.00	0.00
Monterey	# of Records	1435	5134	6680	68602	13719	7631	5375	3391
	% of Ag Records	0.52	1.64	1.91	15.91	2.95	1.66	1.16	0.75
Napa	# of Records	241	150	152	361	51	15	70	90
	% of Ag Records	1.45	0.79	0.68	1.25	0.20	0.05	0.25	0.30
Nevada	# of Records	53	9	14	20	2	0	1	0
	% of Ag Records	7.52	1.74	2.12	3.06	0.28	0.00	0.15	0.00
Orange	# of Records	202	64	316	290	104	168	155	186
	% of Ag Records	1.62	0.72	1.19	1.77	0.29	0.45	0.45	0.56
Placer	# of Records	125	26	47	44	8	15	7	14
	% of Ag Records	3.23	0.75	1.19	1.21	0.21	0.39	0.20	0.36
Plumas	# of Records	1	1	0	0	0	0	0	0
	% of Ag Records	0.65	1.32	0.00	0.00	0.00	0.00	0.00	0.00
Riverside	# of Records	627	1534	3125	2017	1450	2048	345	588
	% of Ag Records	0.88	1.87	4.24	2.99	1.92	2.52	0.45	0.78
Sacramento	# of Records	213	572	261	312	283	71	140	287
	% of Ag Records	1.57	3.44	1.58	1.80	1.76	0.34	0.74	1.32
San Benito	# of Records	3490	193	256	256	377	185	499	490
	% of Ag Records	16.35	0.98	0.96	1.04	1.01	0.51	1.19	1.39
San Bernardino	# of Records	173	273	155	33	23	38	579	30
	% of Ag Records	2.27	2.78	1.47	0.32	0.21	0.30	5.35	0.30
San Diego	# of Records	1447	2070	2026	1649	1929	1827	2147	2182
	% of Ag Records	2.05	2.77	2.49	1.95	2.08	2.04	2.35	1.96
San Francisco	# of Records	1	1	2	0	0	1	0	0
	% of Ag Records	0.57	5.00	5.13	0.00	0.00	4.76	0.00	0.00
San Joaquin	# of Records	1503	1203	1295	1087	4096	2407	2647	1395
	% of Ag Records	1.79	1.66	1.54	1.21	4.20	2.34	2.49	1.26
San Luis Obispo	# of Records	999	748	1288	974	811	1987	2685	3611
	% of Ag Records	1.33	0.86	1.31	1.06	0.83	2.29	2.34	3.01
San Mateo	# of Records	311	228	208	281	502	271	291	328
	% of Ag Records	2.06	1.36	1.06	1.31	2.14	1.28	0.93	1.08
Santa Barbara	# of Records	1294	2306	1694	3052	2365	2944	2503	4418
	% of Ag Records	0.97	1.50	1.12	1.92	1.52	1.81	1.42	2.38
Santa Clara	# of Records	340	171	364	235	366	656	216	347
	% of Ag Records	2.04	1.07	1.65	1.10	1.42	2.59	0.83	1.20
Santa Cruz	# of Records	687	432	10286	4538	627	468	570	621
	% of Ag Records	1.33	0.85	16.26	8.19	1.07	0.87	0.99	1.18
Shasta	# of Records	33	8	10	21	3	8	1	6
	% of Ag Records	2.88	0.59	0.78	2.08	0.22	0.52	0.06	0.29
Sierra	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Siskiyou	# of Records	114	15	95	28	10	52	16	24

	% of Ag Records	2.58	0.45	2.34	0.61	0.17	0.86	0.30	0.40
Solano	# of Records	271	91	125	231	112	103	60	84
	% of Ag Records	1.47	0.50	0.52	1.00	0.48	0.46	0.28	0.41
Sonoma	# of Records	409	790	1070	1211	6214	1073	1405	608
	% of Ag Records	1.59	2.49	3.48	2.93	12.02	2.17	2.84	1.40
Stanislaus	# of Records	943	524	759	2501	4214	3028	2203	3715
	% of Ag Records	1.21	0.59	0.86	2.50	3.75	2.83	1.95	3.33
Sutter	# of Records	398	209	389	349	219	102	147	562
	% of Ag Records	1.93	1.01	1.42	1.40	0.87	0.37	0.50	1.98
Tehama	# of Records	227	59	59	87	59	44	34	37
	% of Ag Records	2.70	0.70	0.63	0.93	0.69	0.36	0.29	0.31
Trinity	# of Records	1	1	0	0	0	0	0	0
	% of Ag Records	1.05	0.45	0.00	0.00	0.00	0.00	0.00	0.00
Tulare	# of Records	1669	2884	3093	3289	3198	3297	3532	3643
	% of Ag Records	0.84	1.49	1.41	1.41	1.28	1.25	1.45	1.54
Tuolumne	# of Records	17	5	0	0	0	0	0	1
	% of Ag Records	2.01	1.14	0.00	0.00	0.00	0.00	0.00	0.11
Ventura	# of Records	1310	1656	1387	1658	2440	2419	1112	214
	% of Ag Records	1.62	2.05	1.61	2.03	2.86	2.79	1.30	0.20
Yolo	# of Records	527	273	520	339	169	173	183	75
	% of Ag Records	1.81	0.95	1.46	0.92	0.45	0.43	0.51	0.24
Yuba	# of Records	74	57	68	80	47	70	29	98
	% of Ag Records	1.32	0.99	1.05	1.27	1.06	0.94	0.37	1.22

Table 3. The number and percent of records reported in the PUR with inconsistent county codes, summarized by county and year.

Cnty_Name	Data	1990	1991	1992	1993	1994	1995	1996	1997
Alameda	# of Records	1797	393	0	0	0	0	0	0
	% of Ag Records	19.99	5.37	0.00	0.00	0.00	0.00	0.00	0.00
Alpine	# of Records	373	6	0	0	0	0	0	0
	% of Ag Records	76.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Amador	# of Records	413	19	0	0	0	0	0	0
	% of Ag Records	20.01	0.61	0.00	0.00	0.00	0.00	0.00	0.00
Butte	# of Records	1749	94	1	0	0	0	0	0
	% of Ag Records	7.61	0.36	0.00	0.00	0.00	0.00	0.00	0.00
Calaveras	# of Records	247	22	0	0	0	0	0	0
	% of Ag Records	28.29	3.73	0.00	0.00	0.00	0.00	0.00	0.00
Colusa	# of Records	783	109	0	0	0	0	0	0
	% of Ag Records	3.80	0.69	0.00	0.00	0.00	0.00	0.00	0.00
Contra Costa	# of Records	853	249	0	0	0	0	0	0
	% of Ag Records	13.26	3.33	0.00	0.00	0.00	0.00	0.00	0.00
Del Norte	# of Records	450	0	0	9	0	0	0	0
	% of Ag Records	12.08	0.00	0.00	0.23	0.00	0.00	0.00	0.00
El Dorado	# of Records	139	42	0	0	4	0	0	0
	% of Ag Records	5.38	1.82	0.00	0.00	0.14	0.00	0.00	0.00
Fresno	# of Records	8377	228	0	0	0	0	0	0
	% of Ag Records	3.03	0.09	0.00	0.00	0.00	0.00	0.00	0.00
Glenn	# of Records	1361	55	0	0	0	0	0	0
	% of Ag Records	6.84	0.26	0.00	0.00	0.00	0.00	0.00	0.00
Humboldt	# of Records	456	3	0	0	0	0	0	0
	% of Ag Records	11.54	0.08	0.00	0.00	0.00	0.00	0.00	0.00
Imperial	# of Records	450	157	0	0	0	0	0	0
	% of Ag Records	0.59	0.16	0.00	0.00	0.00	0.00	0.00	0.00
Inyo	# of Records	112	16	0	0	0	0	0	0
	% of Ag Records	52.34	25.81	0.00	0.00	0.00	0.00	0.00	0.00
Kern	# of Records	3531	127	0	0	0	0	0	0
	% of Ag Records	2.19	0.10	0.00	0.00	0.00	0.00	0.00	0.00
Kings	# of Records	1583	43	0	1	0	0	0	0
	% of Ag Records	3.52	0.08	0.00	0.00	0.00	0.00	0.00	0.00
Lake	# of Records	348	0	0	0	0	0	0	0
	% of Ag Records	5.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lassen	# of Records	60	17	0	0	0	0	0	0
	% of Ag Records	18.87	4.96	0.00	0.00	0.00	0.00	0.00	0.00
Los Angeles	# of Records	743	1162	0	13	1	0	0	0
	% of Ag Records	6.90	12.17	0.00	0.11	0.01	0.00	0.00	0.00
Madera	# of Records	15791	23	0	0	0	0	0	0
	% of Ag Records	26.54	0.04	0.00	0.00	0.00	0.00	0.00	0.00
Marin	# of Records	85	127	0	0	0	0	0	0
	% of Ag Records	11.74	11.16	0.00	0.00	0.00	0.00	0.00	0.00
Mariposa	# of Records	39	28	0	0	0	0	0	0
	% of Ag Records	24.07	26.67	0.00	0.00	0.00	0.00	0.00	0.00
Mendocino	# of Records	191	0	0	0	0	0	0	0

	% of Ag Records	2.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Merced	# of Records	3450	106	0	0	0	0	0	0
	% of Ag Records	4.44	0.13	0.00	0.00	0.00	0.00	0.00	0.00
Modoc	# of Records	154	15	0	0	0	0	0	0
	% of Ag Records	7.70	0.70	0.00	0.00	0.00	0.00	0.00	0.00
Mono	# of Records	96	13	0	3	1	0	0	0
	% of Ag Records	43.24	22.81	0.00	4.23	1.00	0.00	0.00	0.00
Monterey	# of Records	3257	246	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Napa	# of Records	890	25	1	0	0	0	0	74
	% of Ag Records	5.37	0.13	0.00	0.00	0.00	0.00	0.00	0.25
Nevada	# of Records	72	23	0	0	0	0	0	0
	% of Ag Records	10.21	4.45	0.00	0.00	0.00	0.00	0.00	0.00
Orange	# of Records	812	365	0	0	0	0	0	0
	% of Ag Records	6.53	4.10	0.00	0.00	0.00	0.00	0.00	0.00
Placer	# of Records	343	43	0	0	0	0	0	0
	% of Ag Records	8.85	1.24	0.00	0.00	0.00	0.00	0.00	0.00
Plumas	# of Records	104	3	0	0	0	0	0	0
	% of Ag Records	67.53	3.95	0.00	0.00	0.00	0.00	0.00	0.00
Riverside	# of Records	2726	511	2	2	0	0	0	0
	% of Ag Records	3.84	0.62	0.00	0.00	0.00	0.00	0.00	0.00
Sacramento	# of Records	1379	440	0	45	0	0	0	0
	% of Ag Records	10.15	2.65	0.00	0.26	0.00	0.00	0.00	0.00
San Benito	# of Records	7522	17474	0	0	0	0	0	0
	% of Ag Records	35.24	88.30	0.00	0.00	0.00	0.00	0.00	0.00
San Bernardino	# of Records	308	141	0	0	0	0	0	0
	% of Ag Records	4.04	1.44	0.00	0.00	0.00	0.00	0.00	0.00
San Diego	# of Records	1824	881	0	0	0	0	0	0
	% of Ag Records	2.58	1.18	0.00	0.00	0.00	0.00	0.00	0.00
San Francisco	# of Records	16	125	0	0	0	0	0	0
	% of Ag Records	9.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00
San Joaquin	# of Records	2646	247	0	0	0	0	0	0
	% of Ag Records	3.14	0.34	0.00	0.00	0.00	0.00	0.00	0.00
San Luis Obispo	# of Records	882	193	0	0	0	0	0	0
	% of Ag Records	1.17	0.22	0.00	0.00	0.00	0.00	0.00	0.00
San Mateo	# of Records	251	455	0	0	0	0	0	0
	% of Ag Records	1.66	2.71	0.00	0.00	0.00	0.00	0.00	0.00
Santa Barbara	# of Records	7501	213	0	0	3	0	1	0
	% of Ag Records	5.64	0.14	0.00	0.00	0.00	0.00	0.00	0.00
Santa Clara	# of Records	801	583	2	62	0	0	0	0
	% of Ag Records	4.81	3.66	0.01	0.29	0.00	0.00	0.00	0.00
Santa Cruz	# of Records	2271	108	1	0	0	0	0	0
	% of Ag Records	4.39	0.21	0.00	0.00	0.00	0.00	0.00	0.00
Shasta	# of Records	74	221	0	0	0	0	0	0
	% of Ag Records	6.47	16.31	0.00	0.00	0.00	0.00	0.00	0.00
Sierra	# of Records	19	6	0	0	0	0	0	0
	% of Ag Records	40.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Siskiyou	# of Records	191	33	0	0	0	0	0	0

	% of Ag Records	4.32	0.99	0.00	0.00	0.00	0.00	0.00	0.00
Solano	# of Records	2465	164	0	0	0	0	0	0
	% of Ag Records	13.33	0.90	0.00	0.00	0.00	0.00	0.00	0.00
Sonoma	# of Records	486	73	0	0	0	0	0	0
	% of Ag Records	1.88	0.23	0.00	0.00	0.00	0.00	0.00	0.00
Stanislaus	# of Records	3551	179	0	0	0	0	0	0
	% of Ag Records	4.57	0.20	0.00	0.00	0.00	0.00	0.00	0.00
Sutter	# of Records	2425	81	6	0	0	1	0	0
	% of Ag Records	11.73	0.39	0.02	0.00	0.00	0.00	0.00	0.00
Tehama	# of Records	365	95	1	0	0	0	0	0
	% of Ag Records	4.35	1.13	0.01	0.00	0.00	0.00	0.00	0.00
Trinity	# of Records	26	6	0	0	0	0	0	0
	% of Ag Records	27.37	2.68	0.00	0.00	0.00	0.00	0.00	0.00
Tulare	# of Records	7460	76	0	0	0	0	0	0
	% of Ag Records	3.74	0.04	0.00	0.00	0.00	0.00	0.00	0.00
Tuolumne	# of Records	37	34	0	0	0	0	0	0
	% of Ag Records	4.37	7.76	0.00	0.00	0.00	0.00	0.00	0.00
Ventura	# of Records	3464	143	0	2	0	0	0	0
	% of Ag Records	4.28	0.18	0.00	0.00	0.00	0.00	0.00	0.00
Yolo	# of Records	1443	195	0	0	0	0	0	0
	% of Ag Records	4.94	0.68	0.00	0.00	0.00	0.00	0.00	0.00
Yuba	# of Records	342	71	0	0	0	0	0	0
	% of Ag Records	6.10	1.24	0.00	0.00	0.00	0.00	0.00	0.00

Table 4. The number and percent of records reported in the PUR with MTRS values outside the county boundary, summarized by county and year.

CountyName	Data	1990	1991	1992	1993	1994	1995	1996	1997
ALAMEDA	# of Records	3700	47	9	25	42	39	0	0
	% of Ag Records	41.19	0.64	0.14	0.47	0.49	0.55	0.00	0.00
ALPINE	# of Records	486	0	0	0	0	0	0	0
	% of Ag Records	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AMADOR	# of Records	824	1	8	1	0	0	0	0
	% of Ag Records	39.92	0.03	0.47	0.05	0.00	0.00	0.00	0.00
BUTTE	# of Records	1024	150	107	63	76	73	0	0
	% of Ag Records	4.45	0.57	0.38	0.22	0.20	0.21	0.00	0.00
CALAVERAS	# of Records	227	0	1	0	0	0	0	0
	% of Ag Records	26.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00
COLUSA	# of Records	395	1	1	7	2	25	0	0
	% of Ag Records	1.92	0.01	0.00	0.03	0.01	0.09	0.00	0.00
CONTRA COSTA	# of Records	277	50	35	0	0	17	0	0
	% of Ag Records	4.31	0.67	0.42	0.00	0.00	0.18	0.00	0.00
DEL NORTE	# of Records	355	14	51	5	0	0	0	0
	% of Ag Records	9.53	0.47	1.94	0.13	0.00	0.00	0.00	0.00
EL DORADO	# of Records	189	0	0	0	0	0	0	0
	% of Ag Records	7.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FRESNO	# of Records	4082	822	631	481	145	28	0	0
	% of Ag Records	1.48	0.31	0.20	0.14	0.04	0.01	0.00	0.00
GLENN	# of Records	793	10	15	0	31	24	0	0
	% of Ag Records	3.99	0.05	0.08	0.00	0.13	0.10	0.00	0.00
HUMBOLDT	# of Records	321	174	5	15	0	5	0	0
	% of Ag Records	8.12	4.77	0.14	0.58	0.00	0.18	0.00	0.00
IMPERIAL	# of Records	1863	504	124	329	714	1575	0	0
	% of Ag Records	2.46	0.53	0.22	0.44	0.85	1.64	0.00	0.00
INYO	# of Records	167	0	0	0	0	0	0	0
	% of Ag Records	78.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KERN	# of Records	3428	0	1	0	166	72	0	0
	% of Ag Records	2.13	0.00	0.00	0.00	0.10	0.03	0.00	0.00
KINGS	# of Records	592	0	0	3	0	0	20	0
	% of Ag Records	1.32	0.00	0.00	0.00	0.00	0.00	0.03	0.00
LAKE	# of Records	211	7	2	11	58	2	0	0
	% of Ag Records	3.45	0.08	0.03	0.13	1.54	0.02	0.00	0.00
LASSEN	# of Records	65	1	12	0	9	4	0	0
	% of Ag Records	20.44	0.29	2.51	0.00	2.06	1.15	0.00	0.00
LOS ANGELES	# of Records	3195	33	370	769	976	617	0	0
	% of Ag Records	29.65	0.35	3.38	6.38	7.05	5.89	0.00	0.00
MADERA	# of Records	1646	0	0	1	2	0	0	0
	% of Ag Records	2.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MARIN	# of Records	182	2	1	97	250	419	0	0
	% of Ag Records	25.14	0.18	0.08	8.42	18.33	36.12	0.00	0.00
MARIPOSA	# of Records	52	0	7	0	0	0	0	0
	% of Ag Records	32.10	0.00	4.24	0.00	0.00	0.00	0.00	0.00
MENDOCINO	# of Records	223	30	36	5	0	0	0	0

	% of Ag Records	2.83	0.36	0.41	0.05	0.00	0.00	0.00	0.00
MERCED	# of Records	2205	12	5	0	4	8	0	0
	% of Ag Records	2.84	0.01	0.00	0.00	0.00	0.01	0.00	0.00
MODOC	# of Records	241	0	7	3	0	6	0	4
	% of Ag Records	12.04	0.00	0.21	0.09	0.00	0.15	0.00	0.09
MONO	# of Records	150	4	4	0	0	0	0	1
	% of Ag Records	67.57	7.02	5.63	0.00	0.00	0.00	0.00	11.11
MONTEREY	# of Records	8842	393	110	160	289	111	0	0
	% of Ag Records	3.18	0.13	0.03	0.04	0.06	0.02	0.00	0.00
NAPA	# of Records	588	113	159	19	36	41	0	0
	% of Ag Records	3.55	0.60	0.71	0.07	0.14	0.13	0.00	0.00
NEVADA	# of Records	143	9	5	0	1	3	0	0
	% of Ag Records	20.28	1.74	0.76	0.00	0.14	0.51	0.00	0.00
ORANGE	# of Records	2926	174	13373	205	161	238	0	0
	% of Ag Records	23.53	1.96	50.31	1.25	0.44	0.63	0.00	0.00
PLACER	# of Records	269	27	37	47	57	53	18	21
	% of Ag Records	6.94	0.78	0.93	1.29	1.53	1.39	0.52	0.54
PLUMAS	# of Records	116	2	0	0	0	0	0	0
	% of Ag Records	75.32	2.63	0.00	0.00	0.00	0.00	0.00	0.00
RIVERSIDE	# of Records	2992	18	29	33	102	407	0	0
	% of Ag Records	4.21	0.02	0.04	0.05	0.14	0.50	0.00	0.00
SACRAMENTO	# of Records	647	271	298	455	171	446	0	0
	% of Ag Records	4.76	1.63	1.80	2.62	1.06	2.17	0.00	0.00
SAN BENITO	# of Records	929	20	4	7	0	0	0	0
	% of Ag Records	4.35	0.10	0.02	0.03	0.00	0.00	0.00	0.00
SAN BERNARDINO	# of Records	804	146	0	0	0	0	0	0
	% of Ag Records	10.54	1.49	0.00	0.00	0.00	0.00	0.00	0.00
SAN DIEGO	# of Records	6642	773	793	532	606	784	0	0
	% of Ag Records	9.39	1.04	0.97	0.63	0.65	0.88	0.00	0.00
SAN FRANCISCO	# of Records	176	0	0	0	0	0	0	0
	% of Ag Records	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SAN JOAQUIN	# of Records	2334	118	0	10	0	20	0	0
	% of Ag Records	2.77	0.16	0.00	0.01	0.00	0.02	0.00	0.00
SAN LUIS OBISPO	# of Records	11787	927	684	884	726	190	0	0
	% of Ag Records	15.67	1.07	0.70	0.96	0.74	0.22	0.00	0.00
SAN MATEO	# of Records	360	420	514	532	477	289	0	0
	% of Ag Records	2.38	2.51	2.61	2.48	2.04	1.36	0.00	0.00
SANTA BARBARA	# of Records	4404	1317	1866	1386	967	727	0	0
	% of Ag Records	3.31	0.86	1.23	0.87	0.62	0.45	0.00	0.00
SANTA CLARA	# of Records	921	41	25	22	29	0	0	0
	% of Ag Records	5.54	0.26	0.11	0.10	0.11	0.00	0.00	0.00
SANTA CRUZ	# of Records	2658	723	595	0	0	1	0	0
	% of Ag Records	5.14	1.42	0.94	0.00	0.00	0.00	0.00	0.00
SHASTA	# of Records	123	2	6	0	0	0	0	0
	% of Ag Records	10.75	0.15	0.47	0.00	0.00	0.00	0.00	0.00
SIERRA	# of Records	30	1	0	0	2	0	0	0
	% of Ag Records	63.83	25.00	0.00	0.00	13.33	0.00	0.00	0.00
SISKIYOU	# of Records	614	22	68	2	10	10	1	0

	% of Ag Records	13.87	0.66	1.67	0.04	0.17	0.17	0.02	0.00
SOLANO	# of Records	352	96	197	169	111	69	4	0
	% of Ag Records	1.91	0.53	0.82	0.73	0.48	0.31	0.02	0.00
SONOMA	# of Records	882	61	0	0	0	0	0	0
	% of Ag Records	3.42	0.19	0.00	0.00	0.00	0.00	0.00	0.00
STANISLAUS	# of Records	3178	77	175	32	1	50	0	0
	% of Ag Records	4.09	0.09	0.20	0.03	0.00	0.05	0.00	0.00
SUTTER	# of Records	891	302	113	40	59	108	0	0
	% of Ag Records	4.31	1.47	0.41	0.16	0.24	0.39	0.00	0.00
TEHAMA	# of Records	312	61	27	0	0	0	1	0
	% of Ag Records	3.72	0.73	0.29	0.00	0.00	0.00	0.01	0.00
TRINITY	# of Records	45	1	8	6	0	0	0	0
	% of Ag Records	47.37	0.45	4.02	2.11	0.00	0.00	0.00	0.00
TULARE	# of Records	7053	193	136	49	45	48	0	0
	% of Ag Records	3.54	0.10	0.06	0.02	0.02	0.02	0.00	0.00
TUOLUMNE	# of Records	53	4	NA	NA	0	0	NA	0
	% of Ag Records	6.26	0.91	NA	NA	0.00	0.00	NA	0.00
VENTURA	# of Records	4118	492	207	113	268	0	0	0
	% of Ag Records	5.08	0.61	0.24	0.14	0.31	0.00	0.00	0.00
YOLO	# of Records	766	238	297	214	45	57	0	0
	% of Ag Records	2.62	0.83	0.83	0.58	0.12	0.14	0.00	0.00
YUBA	# of Records	302	66	85	0	2	3	0	0
	% of Ag Records	5.38	1.15	1.31	0.00	0.05	0.04	0.00	0.00

NA – refers to the unavailability of the data.

Table 5. The number and percent of records reported in the PUR with a missing location identifier, summarized by county and year.

Cnty_Name	Data	1990	1991	1992	1993	1994	1995	1996	1997
Alameda	# of Records	2	0	0	0	0	0	0	0
	% of Ag Records	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Alpine	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Amador	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Butte	# of Records	23	18	0	0	0	0	0	0
	% of Ag Records	0.10	0.07	0.00	0.00	0.00	0.00	0.00	0.00
Calaveras	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Colusa	# of Records	5	0	2	2	3	7	10	9
	% of Ag Records	0.02	0.00	0.01	0.01	0.01	0.03	0.04	0.04
Contra Costa	# of Records	2	0	1	0	1	0	0	0
	% of Ag Records	0.03	0.00	0.01	0.00	0.01	0.00	0.00	0.00
Del Norte	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
El Dorado	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fresno	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Glenn	# of Records	3	0	0	0	0	0	0	0
	% of Ag Records	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Humboldt	# of Records	8	0	0	0	0	0	0	0
	% of Ag Records	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Imperial	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Inyo	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kern	# of Records	7	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kings	# of Records	1	0	0	0	0	0	1	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lake	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lassen	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Los Angeles	# of Records	1	0	0	0	0	0	0	0
	% of Ag Records	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Madera	# of Records	11	0	0	0	0	0	0	0
	% of Ag Records	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Marin	# of Records	1	0	0	2	0	0	0	0
	% of Ag Records	0.14	0.00	0.00	0.17	0.00	0.00	0.00	0.00
Mariposa	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mendocino	# of Records	0	3	0	0	81	0	0	0
	% of Ag Records	0.00	0.04	0.00	0.00	0.82	0.00	0.00	0.00
Merced	# of Records	9	0	0	0	0	0	0	0
	% of Ag Records	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Modoc	# of Records	0	2	0	0	0	0	0	0
	% of Ag Records	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00
Mono	# of Records	0	0	0	0	0	0	0	1
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	11.11

Monterey	# of Records								
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Napa	# of Records	0	0	1	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Nevada	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Orange	# of Records	0	3	0	0	0	0	0	0
	% of Ag Records	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Placer	# of Records	0	1	0	0	0	0	0	0
	% of Ag Records	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Plumas	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Riverside	# of Records	1	0	0	4	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Sacramento	# of Records	6	0	21	5	0	0	0	0
	% of Ag Records	0.04	0.00	0.13	0.03	0.00	0.00	0.00	0.00
San Benito	# of Records	2	63	0	0	0	0	0	0
	% of Ag Records	0.01	0.32	0.00	0.00	0.00	0.00	0.00	0.00
San Bernardino	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
San Diego	# of Records	3	0	8	0	0	0	0	2
	% of Ag Records	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
San Francisco	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
San Joaquin	# of Records	36	2	0	0	0	0	0	2
	% of Ag Records	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
San Luis Obispo	# of Records	2	3	3	1	4	1	4	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
San Mateo	# of Records	3	0	0	11	0	0	0	0
	% of Ag Records	0.02	0.00	0.00	0.05	0.00	0.00	0.00	0.00
Santa Barbara	# of Records	18	5	0	0	0	0	0	0
	% of Ag Records	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Santa Clara	# of Records	0	0	0	12	38	11	0	0
	% of Ag Records	0.00	0.00	0.00	0.06	0.15	0.04	0.00	0.00
Santa Cruz	# of Records	6	7	3	0	0	0	0	0
	% of Ag Records	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Shasta	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sierra	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Siskiyou	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Solano	# of Records	3	0	0	0	0	0	0	0
	% of Ag Records	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sonoma	# of Records	47	55	0	0	0	0	0	0
	% of Ag Records	0.18	0.17	0.00	0.00	0.00	0.00	0.00	0.00
Stanislaus	# of Records	7	0	0	0	0	0	0	0
	% of Ag Records	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sutter	# of Records	3	0	0	0	0	0	0	0
	% of Ag Records	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tehama	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trinity	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tulare	# of Records	1	0	0	0	0	0	0	0

	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tuolumne	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ventura	# of Records	5	4	2	3	0	0	0	0
	% of Ag Records	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yolo	# of Records	3	1	1	0	0	0	0	0
	% of Ag Records	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yuba	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6. The number and percent of agricultural fields reported in the PUR containing inconsistent MTRS values, summarized by county and year.

Cnty_Name	Data	1990	1991	1992	1993	1994	1995	1996	1997
Alameda	# of Fields	42	19	32	12	0	4	0	0
	% of Ag Fields	3.63	5.16	9.36	4.29	0.00	1.54	0.00	0.00
Alpine	# of Fields	2	0	0	0	0	0	0	0
	% of Ag Fields	1.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Amador	# of Fields	16	10	17	14	9	3	0	0
	% of Ag Fields	2.81	3.92	6.69	6.09	3.30	1.06	0.00	0.00
Butte	# of Fields	276	283	241	11	23	6	2	21
	% of Ag Fields	3.89	6.23	6.10	0.43	0.84	0.23	0.07	0.75
Calaveras	# of Fields	1	4	8	0	1	0	0	0
	% of Ag Fields	0.40	3.88	6.84	0.00	0.57	0.00	0.00	0.00
Colusa	# of Fields	317	4	11	3	7	8	13	11
	% of Ag Fields	4.73	0.14	0.32	0.09	0.19	0.22	0.35	0.31
Contra Costa	# of Fields	77	137	72	1	3	2	1	8
	% of Ag Fields	3.82	9.88	5.79	0.18	0.51	0.35	0.20	1.62
Del Norte	# of Fields	20	6	5	1	4	0	3	0
	% of Ag Fields	4.27	2.27	3.09	0.65	2.90	0.00	4.29	0.00
El Dorado	# of Fields	23	13	27	5	0	0	0	1
	% of Ag Fields	3.58	3.89	7.74	2.62	0.00	0.00	0.00	0.60
Fresno	# of Fields	1272	306	144	59	73	63	66	90
	% of Ag Fields	3.13	1.47	0.68	0.27	0.33	0.28	0.31	0.42
Glenn	# of Fields	153	7	0	1	3	1	5	2
	% of Ag Fields	2.00	0.23	0.00	0.03	0.10	0.03	0.16	0.06
Humboldt	# of Fields	20	9	18	0	0	0	0	1
	% of Ag Fields	3.86	3.56	6.29	0.00	0.00	0.00	0.00	0.43
Imperial	# of Fields	715	22	26	155	169	32	211	174
	% of Ag Fields	4.31	0.33	0.36	2.57	2.82	0.38	3.32	2.80
Inyo	# of Fields	4	0	5	3	0	0	1	0
	% of Ag Fields	4.94	0.00	10.87	6.52	0.00	0.00	3.23	0.00
Kern	# of Fields	745	13	19	13	13	20	17	3
	% of Ag Fields	3.61	0.15	0.21	0.14	0.14	0.20	0.17	0.03
Kings	# of Fields	193	26	12	40	7	18	8	1
	% of Ag Fields	2.41	0.66	0.30	0.97	0.16	0.40	0.17	0.02
Lake	# of Fields	24	27	29	2	13	0	0	0
	% of Ag Fields	1.76	3.90	4.79	0.51	4.04	0.00	0.00	0.00
Lassen	# of Fields	7	3	5	5	5	5	8	6
	% of Ag Fields	4.09	1.97	3.31	4.00	2.59	3.94	4.85	3.03
Los Angeles	# of Fields	144	98	123	81	111	93	7	20
	% of Ag Fields	9.94	15.91	18.72	14.26	14.32	16.76	2.00	5.83
Madera	# of Fields	379	19	31	6	10	6	17	23
	% of Ag Fields	2.93	0.37	0.70	0.14	0.21	0.13	0.37	0.51
Marin	# of Fields	2	2	0	3	1	1	1	0
	% of Ag Fields	1.39	3.57	0.00	6.12	0.74	1.39	2.94	0.00
Mariposa	# of Fields	1	0	4	0	0	0	0	0
	% of Ag Fields	1.67	0.00	8.70	0.00	0.00	0.00	0.00	0.00
Mendocino	# of Fields	65	42	67	7	9	0	4	2

	% of Ag Fields	4.33	5.17	8.54	0.93	1.51	0.00	0.65	0.33
Merced	# of Fields	0	35	48	15	23	8	11	19
	% of Ag Fields	0.00	0.48	0.63	0.20	0.29	0.10	0.14	0.24
Modoc	# of Fields	34	14	13	2	1	0	0	4
	% of Ag Fields	2.98	2.19	1.84	0.40	0.17	0.00	0.00	0.63
Mono	# of Fields	3	5	1	1	4	2	0	0
	% of Ag Fields	3.57	11.63	2.08	1.85	11.11	11.76	0.00	0.00
Monterey	# of Fields	764	101	1	3	3	28	208	1235
	% of Ag Fields	1.76	0.39	0.01	0.02	0.02	0.18	1.20	6.30
Napa	# of Fields	108	108	101	16	4	3	3	6
	% of Ag Fields	3.23	6.59	6.49	1.29	0.34	0.24	0.22	0.45
Nevada	# of Fields	4	13	8	0	0	0	0	0
	% of Ag Fields	1.71	9.63	5.93	0.00	0.00	0.00	0.00	0.00
Orange	# of Fields	137	43	263	35	12	12	10	8
	% of Ag Fields	6.93	5.74	15.77	8.93	1.79	1.75	1.56	1.19
Placer	# of Fields	30	31	30	38	1	4	5	3
	% of Ag Fields	2.56	4.84	4.72	5.36	0.22	0.96	1.23	0.68
Plumas	# of Fields	1	2	1	0	0	0	0	0
	% of Ag Fields	1.49	4.55	3.45	0.00	0.00	0.00	0.00	0.00
Riverside	# of Fields	788	17	57	103	380	44	123	22
	% of Ag Fields	5.27	0.32	1.10	2.23	7.52	0.86	1.74	0.40
Sacramento	# of Fields	118	169	147	159	6	11	37	17
	% of Ag Fields	3.26	5.24	4.47	4.87	0.23	0.40	1.32	0.62
San Benito	# of Fields	152	121	0	1	5	19	9	3
	% of Ag Fields	3.98	4.81	0.00	0.05	0.30	1.20	0.52	0.18
San Bernardino	# of Fields	75	4	15	3	1	0	1	0
	% of Ag Fields	3.74	0.36	2.10	0.39	0.14	0.00	0.14	0.00
San Diego	# of Fields	327	248	4	4	5	2	9	15
	% of Ag Fields	3.68	6.82	0.14	0.14	0.19	0.09	0.38	0.62
San Francisco	# of Fields	2	0	0	0	0	0	0	0
	% of Ag Fields	5.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00
San Joaquin	# of Fields	42	42	6	5	12	7	20	17
	% of Ag Fields	0.16	0.44	0.08	0.07	0.16	0.10	0.26	0.21
San Luis Obispo	# of Fields	759	904	919	732	704	138	7	11
	% of Ag Fields	6.82	13.04	13.41	11.51	10.60	3.42	0.27	0.45
San Mateo	# of Fields	19	24	15	24	35	12	11	12
	% of Ag Fields	0.96	2.66	1.60	2.79	3.78	1.60	4.04	3.14
Santa Barbara	# of Fields	1028	182	954	28	30	28	35	29
	% of Ag Fields	6.04	1.95	10.50	0.37	0.38	0.35	0.42	0.35
Santa Clara	# of Fields	71	99	129	127	74	0	2	24
	% of Ag Fields	1.97	4.43	6.13	6.39	3.70	0.00	0.16	1.97
Santa Cruz	# of Fields	206	176	145	0	2	4	10	2
	% of Ag Fields	3.02	6.83	5.13	0.00	0.13	0.28	0.67	0.16
Shasta	# of Fields	25	16	16	0	1	0	0	2
	% of Ag Fields	4.26	4.08	4.23	0.00	0.30	0.00	0.00	0.53
Sierra	# of Fields	0	0	0	0	0	0	0	0
	% of Ag Fields	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Siskiyou	# of Fields	57	75	29	1	3	3	5	4

	% of Ag Fields	3.21	6.59	2.67	0.12	0.32	0.36	0.50	0.39
Solano	# of Fields	187	220	177	56	11	8	15	23
	% of Ag Fields	2.82	5.66	4.50	1.35	0.39	0.30	0.53	0.89
Sonoma	# of Fields	119	110	3	0	2	4	1	3
	% of Ag Fields	2.48	2.95	0.22	0.00	0.12	0.26	0.06	0.13
Stanislaus	# of Fields	620	455	570	7	42	29	91	39
	% of Ag Fields	3.75	6.02	7.44	0.11	0.67	0.49	1.47	0.66
Sutter	# of Fields	268	295	250	7	15	22	30	7
	% of Ag Fields	3.47	5.86	4.65	0.21	0.45	0.67	0.85	0.20
Tehama	# of Fields	80	96	90	2	0	2	2	6
	% of Ag Fields	3.77	7.17	6.32	0.22	0.00	0.20	0.20	0.56
Trinity	# of Fields	1	0	0	2	0	2	0	6
	% of Ag Fields	2.22	0.00	0.00	2.33	0.00	11.76	0.00	8.33
Tulare	# of Fields	1409	542	243	149	69	30	22	37
	% of Ag Fields	2.85	2.28	1.50	0.93	0.43	0.18	0.14	0.23
Tuolumne	# of Fields	5	3	NA	NA	0	0	NA	7
	% of Ag Fields	2.44	4.05	NA	NA	0.00	0.00	NA	1.84
Ventura	# of Fields	784	416	432	425	321	12	38	3
	% of Ag Fields	7.89	11.96	10.92	11.74	7.78	0.75	2.33	0.18
Yolo	# of Fields	255	280	219	92	19	3	13	19
	% of Ag Fields	2.65	4.60	3.53	1.47	0.45	0.07	0.29	0.46
Yuba	# of Fields	92	83	68	0	4	3	0	1
	% of Ag Fields	5.40	7.85	5.54	0.00	0.98	0.62	0.00	0.10

NA: refers the unavailability of the data.

Table 7. The number and percent of records reported in the PUR containing inconsistent MTRS values for a geographic location, summarized by county and year.

Cnty_Name	Data	1990	1991	1992	1993	1994	1995	1996	1997
Alameda	# of Records	740	987	907	517	0	306	0	0
	% of Ag Records	8.23	13.49	13.69	9.75	0.00	4.34	0.00	0.00
Alpine	# of Records	12	0	0	0	0	0	0	0
	% of Ag Records	2.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Amador	# of Records	120	160	88	99	48	38	0	0
	% of Ag Records	5.81	5.17	5.16	5.13	3.07	1.82	0.00	0.00
Butte	# of Records	2615	4304	4140	135	813	82	50	490
	% of Ag Records	11.37	16.37	14.60	0.46	2.16	0.23	0.15	1.53
Calaveras	# of Records	3	10	33	0	13	0	0	0
	% of Ag Records	0.34	1.69	4.63	0.00	1.29	0.00	0.00	0.00
Colusa	# of Records	2171	30	89	104	157	124	205	126
	% of Ag Records	10.53	0.19	0.41	0.44	0.65	0.46	0.84	0.61
Contra Costa	# of Records	535	1460	970	4	23	28	7	147
	% of Ag Records	8.32	19.53	11.65	0.04	0.24	0.29	0.07	1.15
Del Norte	# of Records	411	83	93	87	62	0	400	0
	% of Ag Records	11.03	2.79	3.53	2.18	1.73	0.00	10.22	0.00
El Dorado	# of Records	397	264	363	28	0	0	0	7
	% of Ag Records	15.36	11.46	13.57	0.97	0.00	0.00	0.00	0.24
Fresno	# of Records	10676	6975	3442	925	1951	1414	1976	2765
	% of Ag Records	3.87	2.61	1.07	0.27	0.54	0.33	0.55	0.73
Glenn	# of Records	976	45	0	69	76	9	80	12
	% of Ag Records	4.91	0.21	0.00	0.34	0.32	0.04	0.35	0.05
Humboldt	# of Records	607	591	564	0	0	0	0	4
	% of Ag Records	15.36	16.20	15.75	0.00	0.00	0.00	0.00	0.16
Imperial	# of Records	6492	769	818	3660	4082	1538	6538	4481
	% of Ag Records	8.57	0.80	1.44	4.89	4.85	1.60	6.92	4.87
Inyo	# of Records	17	0	12	46	0	0	5	0
	% of Ag Records	7.94	0.00	14.12	37.10	0.00	0.00	10.00	0.00
Kern	# of Records	8931	337	353	216	425	685	514	33
	% of Ag Records	5.54	0.27	0.23	0.13	0.24	0.33	0.29	0.02
Kings	# of Records	1415	354	121	916	140	255	247	32
	% of Ag Records	3.15	0.69	0.22	1.47	0.19	0.42	0.34	0.04
Lake	# of Records	266	376	615	87	266	0	0	0
	% of Ag Records	4.35	4.10	8.15	1.03	7.08	0.00	0.00	0.00
Lassen	# of Records	24	9	47	23	19	23	40	29
	% of Ag Records	7.55	2.62	9.81	6.80	4.35	6.61	8.35	4.72
Los Angeles	# of Records	3986	4493	6237	5390	7009	2660	201	2994
	% of Ag Records	36.99	47.06	56.90	44.70	51.13	25.41	1.32	20.76
Madera	# of Records	3860	344	584	141	160	93	438	603
	% of Ag Records	6.49	0.57	0.86	0.20	0.23	0.13	0.68	0.91
Marin	# of Records	22	24	0	75	626	7	31	0
	% of Ag Records	3.04	2.11	0.00	6.51	45.89	0.60	6.95	0.00
Mariposa	# of Records	2	0	29	0	0	0	0	0
	% of Ag Records	1.23	0.00	17.58	0.00	0.00	0.00	0.00	0.00
Mendocino	# of Records	865	719	1364	62	186	0	25	69

	% of Ag Records	10.99	8.73	15.45	0.64	1.88	0.00	0.25	0.65
Merced	# of Records	0	818	859	234	912	153	226	428
	% of Ag Records	0.00	1.00	0.85	0.21	0.80	0.11	0.18	0.35
Modoc	# of Records	110	65	79	26	3	0	0	92
	% of Ag Records	5.50	3.02	2.35	0.74	0.12	0.00	0.00	2.01
Mono	# of Records	33	16	3	2	27	7	0	0
	% of Ag Records	14.86	28.07	3.19	2.82	27.00	26.92	0.00	0.00
Monterey	# of Records	9198	1053	53	42	155	603	7647	43648
	% of Ag Records	3.30	0.34	0.02	0.01	0.03	0.13	1.66	9.65
Napa	# of Records	1285	2309	2242	2501	128	98	66	133
	% of Ag Records	7.76	12.21	10.02	8.65	0.50	0.31	0.24	0.45
Nevada	# of Records	46	114	71	0	0	0	0	0
	% of Ag Records	6.52	22.05	10.77	0.00	0.00	0.00	0.00	0.00
Orange	# of Records	2280	285	15363	2283	2521	24	20	17
	% of Ag Records	18.34	3.20	57.80	13.94	6.92	0.06	0.06	0.05
Placer	# of Records	281	328	469	526	3	77	94	79
	% of Ag Records	7.25	9.46	11.85	14.49	0.08	2.01	2.69	2.05
Plumas	# of Records	3	7	2	0	0	0	0	0
	% of Ag Records	1.95	9.21	5.71	0.00	0.00	0.00	0.00	0.00
Riverside	# of Records	8210	981	958	2186	9469	845	2142	426
	% of Ag Records	11.55	1.19	1.30	3.25	12.54	1.04	2.78	0.57
Sacramento	# of Records	1457	1595	1326	1702	47	109	476	226
	% of Ag Records	10.72	9.59	8.03	9.81	0.29	0.53	2.50	1.04
San Benito	# of Records	1486	1169	0	9	127	480	156	49
	% of Ag Records	6.96	5.91	0.00	0.04	0.34	1.33	0.37	0.14
San Bernardino	# of Records	548	49	326	158	6	0	34	0
	% of Ag Records	7.19	0.50	3.10	1.52	0.06	0.00	0.31	0.00
San Diego	# of Records	8885	18683	152	37	376	163	463	1399
	% of Ag Records	12.57	25.02	0.19	0.04	0.41	0.18	0.51	1.26
San Francisco	# of Records	32	0	0	0	0	0	0	0
	% of Ag Records	18.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00
San Joaquin	# of Records	976	976	78	133	142	133	586	281
	% of Ag Records	1.16	1.35	0.09	0.15	0.15	0.13	0.55	0.25
San Luis Obispo	# of Records	11626	28479	36016	25697	27482	3166	1169	1552
	% of Ag Records	15.46	32.80	36.65	27.84	28.02	3.65	1.02	1.29
San Mateo	# of Records	471	809	3865	6349	4349	569	756	1610
	% of Ag Records	3.12	4.83	19.61	29.64	18.56	2.68	2.41	5.30
Santa Barbara	# of Records	23011	41428	36848	5500	5260	5664	2152	4605
	% of Ag Records	17.32	26.91	24.38	3.46	3.39	3.49	1.22	2.48
Santa Clara	# of Records	545	1644	3299	3064	1458	0	11	999
	% of Ag Records	3.28	10.33	14.99	14.39	5.65	0.00	0.04	3.46
Santa Cruz	# of Records	4307	9652	7063	0	74	292	163	316
	% of Ag Records	8.33	19.02	11.17	0.00	0.13	0.54	0.28	0.60
Shasta	# of Records	92	149	71	0	3	0	0	0
	% of Ag Records	8.04	11.00	5.54	0.00	0.22	0.00	0.00	0.00
Sierra	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Siskiyou	# of Records	231	572	171	7	14	43	56	56

	% of Ag Records	5.22	17.09	4.21	0.15	0.24	0.71	1.04	0.94
Solano	# of Records	1410	2448	3996	490	211	101	103	227
	% of Ag Records	7.63	13.47	16.69	2.13	0.91	0.45	0.47	1.11
Sonoma	# of Records	2382	1445	48	0	66	128	36	49
	% of Ag Records	9.24	4.55	0.16	0.00	0.13	0.26	0.07	0.11
Stanislaus	# of Records	6023	8035	10483	341	22350	12837	6025	15462
	% of Ag Records	7.74	9.04	11.92	0.34	19.87	11.98	5.34	13.85
Sutter	# of Records	1501	2830	2274	118	223	305	433	121
	% of Ag Records	7.26	13.74	8.33	0.47	0.89	1.11	1.47	0.43
Tehama	# of Records	1036	1263	1175	16	0	149	17	272
	% of Ag Records	12.34	15.06	12.51	0.17	0.00	1.20	0.14	2.26
Trinity	# of Records	5	0	0	18	0	33	0	26
	% of Ag Records	5.26	0.00	0.00	6.34	0.00	33.33	0.00	15.12
Tulare	# of Records	11049	8180	4121	2369	1255	474	372	715
	% of Ag Records	5.55	4.24	1.87	1.02	0.50	0.18	0.15	0.30
Tuolumne	# of Records	23	6	0	0	0	0	0	33
	% of Ag Records	2.72	1.37	0.00	0.00	0.00	0.00	0.00	3.66
Ventura	# of Records	20999	31096	30412	30583	16735	370	5716	11
	% of Ag Records	25.92	38.42	35.20	37.52	19.65	0.43	6.68	0.17
Yolo	# of Records	1773	3113	2403	834	248	72	174	231
	% of Ag Records	6.07	10.83	6.73	2.25	0.67	0.18	0.49	0.75
Yuba	# of Records	836	1031	856	0	97	40	0	14
	% of Ag Records	14.90	17.95	13.16	0.00	2.19	0.54	0.00	0.17

Table 8. The number and percent of agricultural fields with inconsistent acres planted reported in the PUR, summarized by county and year.

Cnty_Name	Data	1990	1991	1992	1993	1994	1995	1996	1997
Alameda	# of Fields	35	45	31	23	13	9	5	5
	% of Ag Fields	3.03	12.23	9.06	8.21	3.53	3.46	3.09	1.53
Alpine	# of Fields	4	0	0	0	0	0	0	0
	% of Ag Fields	2.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Amador	# of Fields	14	22	34	21	12	11	2	3
	% of Ag Fields	2.46	8.63	13.39	9.13	4.40	3.89	1.13	1.61
Butte	# of Fields	456	505	520	72	60	50	40	58
	% of Ag Fields	6.42	11.11	13.17	2.83	2.19	1.91	1.44	2.06
Calaveras	# of Fields	5	6	9	0	1	3	0	1
	% of Ag Fields	2.02	5.83	7.69	0.00	0.57	2.65	0.00	0.70
Colusa	# of Fields	409	20	47	41	56	59	57	50
	% of Ag Fields	6.10	0.72	1.36	1.19	1.53	1.62	1.56	1.41
Contra Costa	# of Fields	94	136	119	9	15	27	12	29
	% of Ag Fields	4.66	9.81	9.57	1.59	2.53	4.79	2.42	5.86
Del Norte	# of Fields	26	20	12	16	5	6	5	4
	% of Ag Fields	5.56	7.58	7.41	10.46	3.62	8.57	7.14	5.19
El Dorado	# of Fields	39	45	47	13	10	3	1	2
	% of Ag Fields	6.07	13.47	13.47	6.81	5.78	2.01	0.63	1.19
Fresno	# of Fields	1641	494	1408	1216	1010	1001	832	858
	% of Ag Fields	4.04	2.37	6.62	5.52	4.59	4.43	3.92	3.96
Glenn	# of Fields	230	19	16	39	22	37	62	37
	% of Ag Fields	3.01	0.62	0.58	1.34	0.71	1.26	1.98	1.20
Humboldt	# of Fields	15	7	40	1	0	2	2	2
	% of Ag Fields	2.90	2.77	13.99	0.57	0.00	1.01	0.83	0.85
Imperial	# of Fields	1031	240	114	244	199	135	204	272
	% of Ag Fields	6.21	3.65	1.58	4.05	3.32	1.59	3.21	4.37
Inyo	# of Fields	3	2	8	2	0	0	0	0
	% of Ag Fields	3.70	6.67	17.39	4.35	0.00	0.00	0.00	0.00
Kern	# of Fields	626	43	64	44	41	35	39	34
	% of Ag Fields	3.03	0.50	0.72	0.47	0.44	0.35	0.39	0.35
Kings	# of Fields	361	33	31	62	56	37	28	37
	% of Ag Fields	4.51	0.84	0.77	1.51	1.32	0.82	0.60	0.82
Lake	# of Fields	33	54	65	12	29	6	4	17
	% of Ag Fields	2.42	7.80	10.73	3.05	9.01	1.84	1.49	4.68
Lassen	# of Fields	11	14	12	17	12	11	19	10
	% of Ag Fields	6.43	9.21	7.95	13.60	6.22	8.66	11.52	5.05
Los Angeles	# of Fields	89	127	118	117	94	50	33	31
	% of Ag Fields	6.14	20.62	17.96	20.60	12.13	9.01	9.43	9.04
Madera	# of Fields	549	25	50	23	38	15	23	18
	% of Ag Fields	4.24	0.49	1.12	0.52	0.81	0.33	0.50	0.40
Marin	# of Fields	3	5	4	3	9	6	2	0
	% of Ag Fields	2.08	8.93	9.09	6.12	6.67	8.33	5.88	0.00
Mariposa	# of Fields	0	3	2	0	1	1	0	1
	% of Ag Fields	0.00	9.38	4.35	0.00	1.69	1.69	0.00	0.50
Mendocino	# of Fields	70	79	108	12	3	5	7	9
	% of Ag Fields	4.66	9.73	13.76	1.60	0.50	0.89	1.14	1.50
Merced	# of Fields	0	137	109	130	105	91	82	128
	% of Ag Fields	0.00	1.89	1.43	1.74	1.34	1.15	1.04	1.63
Modoc	# of Fields	31	18	27	7	9	11	4	9
	% of Ag Fields	2.71	2.82	3.81	1.42	1.52	2.55	0.68	1.42
Mono	# of Fields	2	5	1	4	2	2	0	0
	% of Ag Fields	2.38	11.63	2.08	7.41	5.56	11.76	0.00	0.00

Monterey	# of Fields	1680	2763	5208	4693	4819	5447	5832	5847
	% of Ag Fields	3.86	10.65	32.90	29.68	31.03	35.26	33.69	29.82
Napa	# of Fields	156	258	282	88	34	68	55	48
	% of Ag Fields	4.67	15.73	18.12	7.11	2.86	5.44	4.07	3.60
Nevada	# of Fields	7	18	22	0	0	0	0	0
	% of Ag Fields	2.99	13.33	16.30	0.00	0.00	0.00	0.00	0.00
Orange	# of Fields	179	137	264	37	55	58	50	31
	% of Ag Fields	9.06	18.29	15.83	9.44	8.22	8.48	7.82	4.60
Placer	# of Fields	58	85	66	68	2	2	5	7
	% of Ag Fields	4.96	13.26	10.39	9.59	0.43	0.48	1.23	1.58
Plumas	# of Fields	3	2	0	2	0	0	0	0
	% of Ag Fields	4.48	4.55	0.00	8.00	0.00	0.00	0.00	0.00
Riverside	# of Fields	968	206	461	587	690	570	871	888
	% of Ag Fields	6.47	3.89	8.93	12.69	13.66	11.18	12.30	16.05
Sacramento	# of Fields	231	298	221	295	25	40	38	17
	% of Ag Fields	6.37	9.23	6.72	9.04	0.95	1.46	1.35	0.62
San Benito	# of Fields	321	380	639	566	696	708	754	698
	% of Ag Fields	8.41	15.12	29.18	29.68	41.33	44.84	43.61	41.16
San Bernardino	# of Fields	75	20	47	49	35	36	11	20
	% of Ag Fields	3.74	1.79	6.58	6.40	4.73	4.83	1.58	2.77
San Diego	# of Fields	380	394	96	57	91	95	47	72
	% of Ag Fields	4.28	10.83	3.27	2.01	3.42	4.12	1.97	2.97
San Francisco	# of Fields	1	1	2	0	0	1	0	0
	% of Ag Fields	2.56	12.50	16.67	0.00	0.00	14.29	0.00	0.00
San Joaquin	# of Fields	640	95	42	108	75	73	92	83
	% of Ag Fields	2.39	0.99	0.56	1.51	1.03	1.00	1.21	1.00
San Luis Obispo	# of Fields	1419	1635	1617	1621	1777	1552	1598	1686
	% of Ag Fields	12.75	23.59	23.60	25.49	26.75	38.49	61.87	69.53
San Mateo	# of Fields	118	208	338	115	78	75	26	32
	% of Ag Fields	5.94	23.09	36.03	13.36	8.42	9.99	9.56	8.38
Santa Barbara	# of Fields	2220	2706	2405	2665	2659	2550	2929	3066
	% of Ag Fields	13.05	29.06	26.46	35.68	33.70	32.26	35.23	37.44
Santa Clara	# of Fields	222	333	350	354	209	17	55	63
	% of Ag Fields	6.18	14.91	16.62	17.82	10.44	1.58	4.32	5.16
Santa Cruz	# of Fields	382	573	355	103	136	93	37	32
	% of Ag Fields	5.59	22.25	12.55	6.53	8.79	6.49	2.48	2.59
Shasta	# of Fields	19	27	20	5	3	1	2	7
	% of Ag Fields	3.24	6.89	5.29	1.92	0.90	0.38	0.63	1.85
Sierra	# of Fields	0	0	0	0	0	0	0	0
	% of Ag Fields	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Siskiyou	# of Fields	37	16	52	12	8	3	18	10
	% of Ag Fields	2.08	1.41	4.78	1.41	0.85	0.36	1.79	0.98
Solano	# of Fields	179	253	261	62	58	43	45	44
	% of Ag Fields	2.70	6.51	6.63	1.49	2.04	1.62	1.60	1.71
Sonoma	# of Fields	212	225	74	64	84	59	73	145
	% of Ag Fields	4.41	6.04	5.33	4.42	5.22	3.80	4.53	6.46
Stanislaus	# of Fields	1266	1092	1000	88	162	70	240	263
	% of Ag Fields	7.67	14.44	13.05	1.44	2.57	1.17	3.88	4.43
Sutter	# of Fields	360	485	404	35	39	92	85	3
	% of Ag Fields	4.66	9.63	7.52	1.05	1.17	2.81	2.40	0.09
Tehama	# of Fields	100	158	121	18	19	27	19	24
	% of Ag Fields	4.71	11.80	8.50	1.96	2.37	2.72	1.86	2.22
Trinity	# of Fields	0	5	7	5	1	0	2	3
	% of Ag Fields	0.00	6.02	10.45	5.81	2.27	0.00	5.88	4.17
Tulare	# of Fields	2240	1107	516	294	320	316	138	115

	% of Ag Fields	4.52	4.65	3.18	1.84	2.01	1.92	0.85	0.71
Tuolumne	# of Fields	10	5	NA	NA	0	0	NA	5
	% of Ag Fields	4.88	6.76	NA	NA	0.00	0.00	NA	1.31
Ventura	# of Fields	2137	1625	1532	1607	1869	1435	1214	111
	% of Ag Fields	21.49	46.74	38.73	44.38	45.28	89.63	74.30	6.83
Yolo	# of Fields	295	420	330	110	33	33	27	33
	% of Ag Fields	3.07	6.91	5.32	1.75	0.79	0.79	0.61	0.80
Yuba	# of Fields	103	137	130	13	21	18	41	21
	% of Ag Fields	6.04	12.96	10.59	2.81	5.12	3.71	4.30	2.20

NA: refers to the unavailability of the data.

Table 9. The number and percent of records reported in the PUR with inconsistent acres planted, summarized by county and year.

Cnty_Name	Data	1990	1991	1992	1993	1994	1995	1996	1997
Alameda	# of Records	1276	3268	2278	3551	2538	1241	1184	139
	% of Ag Records	14.19	44.67	34.38	66.94	29.47	17.60	24.64	1.77
Alpine	# of Records	21	0	0	0	0	0	0	0
	% of Ag Records	4.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Amador	# of Records	97	2263	340	275	116	121	11	128
	% of Ag Records	4.70	73.09	19.93	14.25	7.42	5.80	0.63	4.63
Butte	# of Records	3496	5894	7472	1182	1097	1255	614	747
	% of Ag Records	15.21	22.42	26.34	4.05	2.91	3.59	1.79	2.32
Calaveras	# of Records	32	28	65	0	19	9	0	2
	% of Ag Records	3.67	4.75	9.13	0.00	1.88	1.11	0.00	0.23
Colusa	# of Records	2238	528	492	905	773	879	678	337
	% of Ag Records	10.85	3.33	2.28	3.86	3.22	3.25	2.78	1.62
Contra Costa	# of Records	677	1899	1200	111	248	408	273	673
	% of Ag Records	10.52	25.40	14.41	1.09	2.63	4.24	2.65	5.25
Del Norte	# of Records	388	345	191	470	84	274	440	302
	% of Ag Records	10.41	11.59	7.26	11.80	2.35	7.39	11.24	7.08
El Dorado	# of Records	333	506	463	276	293	118	13	19
	% of Ag Records	12.89	21.96	17.31	9.55	10.15	5.35	0.51	0.65
Fresno	# of Records	13357	9371	32453	28793	23954	26102	21200	20475
	% of Ag Records	4.84	3.50	10.13	8.38	6.69	6.14	5.95	5.43
Glenn	# of Records	1338	97	156	625	260	336	858	628
	% of Ag Records	6.72	0.46	0.78	3.09	1.09	1.43	3.78	2.67
Humboldt	# of Records	166	93	1327	5	0	45	10	9
	% of Ag Records	4.20	2.55	37.06	0.19	0.00	1.62	0.35	0.37
Imperial	# of Records	7946	6964	2483	6515	5883	3952	5475	6612
	% of Ag Records	10.49	7.27	4.38	8.70	6.99	4.11	5.80	7.18
Inyo	# of Records	8	17	33	12	0	0	0	0
	% of Ag Records	3.74	27.42	38.82	9.68	0.00	0.00	0.00	0.00
Kern	# of Records	7009	578	1088	861	1095	1330	636	800
	% of Ag Records	4.35	0.46	0.71	0.54	0.63	0.64	0.36	0.47
Kings	# of Records	2243	671	795	1314	1275	656	678	1264
	% of Ag Records	4.99	1.30	1.42	2.11	1.76	1.09	0.92	1.57
Lake	# of Records	374	3502	1464	451	575	358	62	695
	% of Ag Records	6.12	38.15	19.39	5.32	15.30	4.24	1.36	7.71
Lassen	# of Records	33	94	82	68	62	46	107	51
	% of Ag Records	10.38	27.41	17.12	20.12	14.19	13.22	22.34	8.31
Los Angeles	# of Records	1138	3456	3954	3913	3234	3093	1736	3367
	% of Ag Records	10.56	36.20	36.07	32.45	23.59	29.55	11.41	23.35
Madera	# of Records	5856	483	1320	555	690	254	586	418
	% of Ag Records	9.84	0.80	1.94	0.78	0.99	0.34	0.91	0.63
Marin	# of Records	18	636	790	71	114	225	32	0
	% of Ag Records	2.49	55.89	64.49	6.16	8.36	19.40	7.17	0.00
Mariposa	# of Records	0	19	17	0	3	4	0	3
	% of Ag Records	0.00	18.10	10.30	0.00	1.36	1.42	0.00	0.38

Mendocino	# of Records	819	1400	2215	118	52	257	71	154
	% of Ag Records	10.41	17.00	25.08	1.22	0.52	2.37	0.72	1.46
Merced	# of Records	0	2966	2531	2976	2302	2617	1619	3352
	% of Ag Records	0.00	3.62	2.51	2.71	2.02	1.82	1.29	2.71
Modoc	# of Records	100	110	230	136	63	181	24	137
	% of Ag Records	5.00	5.11	6.86	3.88	2.62	4.39	0.61	3.00
Mono	# of Records	30	14	2	13	13	7	0	0
	% of Ag Records	13.51	24.56	2.13	18.31	13.00	26.92	0.00	0.00
Monterey	# of Records	18668	55924	132803	155248	175550	179034	175122	166202
	% of Ag Records	6.71	17.83	38.04	36.01	37.74	39.05	37.91	36.73
Napa	# of Records	1823	5719	6687	3323	1101	2446	1041	1411
	% of Ag Records	11.00	30.25	29.89	11.49	4.29	7.82	3.73	4.73
Nevada	# of Records	27	130	186	0	0	0	0	0
	% of Ag Records	3.83	25.15	28.22	0.00	0.00	0.00	0.00	0.00
Orange	# of Records	1781	2319	13132	1710	5540	4191	9891	3251
	% of Ag Records	14.32	26.08	49.41	10.44	15.20	11.16	28.82	9.76
Placer	# of Records	285	650	611	545	27	66	368	102
	% of Ag Records	7.36	18.74	15.43	15.01	0.73	1.73	10.55	2.64
Plumas	# of Records	12	10	0	4	0	0	0	0
	% of Ag Records	7.79	13.16	0.00	7.69	0.00	0.00	0.00	0.00
Riverside	# of Records	9666	5168	9143	10498	13658	11692	16591	16568
	% of Ag Records	13.60	6.29	12.41	15.59	18.09	14.38	21.53	22.00
Sacramento	# of Records	2335	3999	2534	3332	653	602	666	498
	% of Ag Records	17.18	24.04	15.34	19.20	4.06	2.92	3.50	2.29
San Benito	# of Records	4206	6477	13361	11963	22650	22239	23818	17246
	% of Ag Records	19.70	32.73	50.13	48.73	60.95	61.48	56.72	48.81
San Bernardino	# of Records	778	424	1135	1661	1103	1590	319	576
	% of Ag Records	10.20	4.32	10.80	15.98	10.29	12.68	2.95	5.75
San Diego	# of Records	6671	24605	5288	3256	13890	16215	6334	7831
	% of Ag Records	9.44	32.95	6.50	3.86	15.00	18.11	6.93	7.04
San Francisco	# of Records	22	11	29	0	0	2	0	0
	% of Ag Records	12.50	55.00	74.36	0.00	0.00	9.52	0.00	0.00
San Joaquin	# of Records	4876	1640	589	2378	1825	1759	1828	1924
	% of Ag Records	5.79	2.26	0.70	2.65	1.87	1.71	1.72	1.74
San Luis Obispo	# of Records	19537	41096	45580	41767	48973	45112	67317	71655
	% of Ag Records	25.98	47.33	46.38	45.24	49.93	51.96	58.65	59.79
San Mateo	# of Records	4569	10000	12802	11622	10880	8052	7985	7077
	% of Ag Records	30.24	59.65	64.95	54.26	46.44	37.89	25.50	23.28
Santa Barbara	# of Records	36430	71957	66980	77689	68233	75454	87545	96647
	% of Ag Records	27.42	46.74	44.32	48.83	43.92	46.45	49.80	51.98
Santa Clara	# of Records	2452	5772	6768	8136	4971	1066	1753	3325
	% of Ag Records	14.74	36.27	30.75	38.20	19.25	4.21	6.75	11.51
Santa Cruz	# of Records	7442	24412	17027	8530	9785	5223	3949	2350
	% of Ag Records	14.39	48.11	26.92	15.39	16.71	9.67	6.84	4.48
Shasta	# of Records	85	169	95	43	32	5	15	72
	% of Ag Records	7.43	12.47	7.41	4.27	2.33	0.32	0.87	3.52
Sierra	# of Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Siskiyou	# of Records	123	106	376	183	177	50	176	77
	% of Ag Records	2.78	3.17	9.25	4.02	2.99	0.83	3.28	1.29
Solano	# of Records	1045	2681	2422	572	679	507	360	682
	% of Ag Records	5.65	14.75	10.12	2.48	2.92	2.28	1.65	3.34
Sonoma	# of Records	2820	3029	4380	7952	3794	2526	5655	8061
	% of Ag Records	10.94	9.54	14.26	19.21	7.34	5.10	11.42	18.51
Stanislaus	# of Records	12559	19223	19299	2454	7171	6573	7343	16677
	% of Ag Records	16.15	21.62	21.95	2.45	6.38	6.13	6.50	14.94
Sutter	# of Records	2087	3721	3346	574	550	1071	1278	585
	% of Ag Records	10.10	18.06	12.25	2.30	2.19	3.91	4.33	2.06
Tehama	# of Records	1033	1957	1320	316	365	585	216	738
	% of Ag Records	12.30	23.33	14.05	3.39	4.28	4.72	1.82	6.12
Trinity	# of Records	0	29	28	23	5	0	6	17
	% of Ag Records	0.00	12.95	14.07	8.10	3.14	0.00	4.00	9.88
Tulare	# of Records	17752	15711	10839	7535	7526	8448	6151	2483
	% of Ag Records	8.91	8.14	4.93	3.23	3.02	3.19	2.53	1.05
Tuolumne	# of Records	50	33	NA	NA	0	0	NA	21
	% of Ag Records	5.91	7.53	NA	NA	0.00	0.00	NA	2.33
Ventura	# of Records	35903	56210	56007	52867	53852	70357	62068	3156
	% of Ag Records	44.32	69.46	64.82	64.85	63.22	81.02	72.53	48.22
Yolo	# of Records	2138	4022	3545	1137	370	432	340	363
	% of Ag Records	7.32	13.99	9.93	3.07	0.99	1.08	0.95	1.18
Yuba	# of Records	846	1525	1474	270	550	402	627	478
	% of Ag Records	15.08	26.54	22.65	4.30	12.41	5.39	8.10	5.94

NA: refers the unavailability of the data.

Table 10. The number and percent of records reported in the PUR with acres treated greater than acres planted, summarized by county and year.

Cnty_Name	Data	1990	1991	1992	1993	1994	1995	1996	1997
Alameda	# Records	544	1370	470	1120	90	21	1	0
	% of Ag Records	6.05	18.73	7.09	21.11	1.04	0.30	0.02	0.00
Alpine	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Amador	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Butte	# Records	19	181	100	3	0	0	1	0
	% of Ag Records	0.08	0.69	0.35	0.01	0.00	0.00	0.00	0.00
Calaveras	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Colusa	# Records	0	3	0	0	0	0	0	0
	% of Ag Records	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Contra Costa	# Records	25	570	100	0	0	0	0	0
	% of Ag Records	0.39	7.62	1.20	0.00	0.00	0.00	0.00	0.00
Del Norte	# Records	72	83	0	1	0	0	0	0
	% of Ag Records	1.93	2.79	0.00	0.03	0.00	0.00	0.00	0.00
El Dorado	# Records	4	11	0	0	103	73	0	0
	% of Ag Records	0.15	0.48	0.00	0.00	3.57	3.31	0.00	0.00
Fresno	# Records	49	27	0	0	0	0	2	0
	% of Ag Records	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Glenn	# Records	1	1	0	0	0	0	0	0
	% of Ag Records	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Humboldt	# Records	12	12	9	0	0	32	0	0
	% of Ag Records	0.30	0.33	0.25	0.00	0.00	1.15	0.00	0.00
Imperial	# Records	0	0	3	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Inyo	# Records	0	2	0	0	0	0	0	0
	% of Ag Records	0.00	3.23	0.00	0.00	0.00	0.00	0.00	0.00
Kern	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kings	# Records	2	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lake	# Records	0	4	7	0	1	0	0	0
	% of Ag Records	0.00	0.04	0.09	0.00	0.03	0.00	0.00	0.00
Lassen	# Records	2	23	12	0	1	0	1	0
	% of Ag Records	0.63	6.71	2.51	0.00	0.23	0.00	0.21	0.00
Los Angeles	# Records	217	768	1812	1126	406	220	69	448
	% of Ag Records	2.01	8.04	16.53	9.34	2.96	2.10	0.45	3.11
Madera	# Records	237	0	0	0	0	0	0	0
	% of Ag Records	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Marin	# Records	0	447	81	0	155	108	32	0
	% of Ag Records	0.00	39.28	6.61	0.00	11.36	9.31	7.17	0.00
Mariposa	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mendocino	# Records	6	198	54	0	0	0	0	1

	% of Ag Records	0.08	2.40	0.61	0.00	0.00	0.00	0.00	0.01
Merced	# Records	0	17	2	5	0	12	11	13
	% of Ag Records	0.00	0.02	0.00	0.00	0.00	0.01	0.01	0.01
Modoc	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mono	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Monterey	# Records	1313	1034	0	0	4320	1813	5674	3784
	% of Ag Records	0.47	0.33	0.00	0.00	0.93	0.40	1.23	0.84
Napa	# Records	1	60	9	0	0	0	0	0
	% of Ag Records	0.01	0.32	0.04	0.00	0.00	0.00	0.00	0.00
Nevada	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Orange	# Records	468	462	8963	0	0	0	4	0
	% of Ag Records	3.76	5.20	33.72	0.00	0.00	0.00	0.01	0.00
Placer	# Records	29	20	9	47	0	0	7	4
	% of Ag Records	0.75	0.58	0.23	1.29	0.00	0.00	0.20	0.10
Plumas	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Riverside	# Records	213	0	0	0	0	0	79	1
	% of Ag Records	0.30	0.00	0.00	0.00	0.00	0.00	0.10	0.00
Sacramento	# Records	5	1	45	54	5	0	0	0
	% of Ag Records	0.04	0.01	0.27	0.31	0.03	0.00	0.00	0.00
San Benito	# Records	354	1398	0	0	1655	1	13232	8990
	% of Ag Records	1.66	7.06	0.00	0.00	4.45	0.00	31.51	25.44
San Bernardino	# Records	217	0	0	0	0	0	0	0
	% of Ag Records	2.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00
San Diego	# Records	1160	3188	0	0	261	73	1136	324
	% of Ag Records	1.64	4.27	0.00	0.00	0.28	0.08	1.24	0.29
San Francisco	# Records	5	0	9	0	0	0	0	0
	% of Ag Records	2.84	0.00	23.08	0.00	0.00	0.00	0.00	0.00
San Joaquin	# Records	0	689	5	0	2	5	3	2
	% of Ag Records	0.00	0.95	0.01	0.00	0.00	0.00	0.00	0.00
San Luis Obispo	# Records	230	1729	2025	42	1	0	976	1129
	% of Ag Records	0.31	1.99	2.06	0.05	0.00	0.00	0.85	0.94
San Mateo	# Records	2013	6063	4057	2515	3931	1990	911	288
	% of Ag Records	13.32	36.17	20.58	11.74	16.78	9.36	2.91	0.95
Santa Barbara	# Records	948	3491	1960	3998	2946	2826	4781	5087
	% of Ag Records	0.71	2.27	1.30	2.51	1.90	1.74	2.72	2.74
Santa Clara	# Records	109	3832	365	765	499	343	46	155
	% of Ag Records	0.66	24.08	1.66	3.59	1.93	1.35	0.18	0.54
Santa Cruz	# Records	401	4444	1537	0	1722	319	177	368
	% of Ag Records	0.78	8.76	2.43	0.00	2.94	0.59	0.31	0.70
Shasta	# Records	0	2	0	0	0	0	0	0
	% of Ag Records	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00
Sierra	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Siskiyou	# Records	0	0	6	0	0	0	0	0

	% of Ag Records	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00
Solano	# Records	0	1	8	0	0	0	0	0
	% of Ag Records	0.00	0.01	0.03	0.00	0.00	0.00	0.00	0.00
Sonoma	# Records	165	218	0	0	669	676	1166	100
	% of Ag Records	0.64	0.69	0.00	0.00	1.29	1.37	2.35	0.23
Stanislaus	# Records	65	39	49	0	0	0	0	0
	% of Ag Records	0.08	0.04	0.06	0.00	0.00	0.00	0.00	0.00
Sutter	# Records	4	12	0	0	0	0	0	0
	% of Ag Records	0.02	0.06	0.00	0.00	0.00	0.00	0.00	0.00
Tehama	# Records	0	10	4	0	0	0	0	0
	% of Ag Records	0.00	0.12	0.04	0.00	0.00	0.00	0.00	0.00
Trinity	# Records	0	0	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tulare	# Records	8	4	0	0	0	0	0	0
	% of Ag Records	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tuolumne	# Records	0	15	NA	NA	0	0	NA	0
	% of Ag Records	0.00	3.42	NA	NA	0.00	0.00	NA	0.00
Ventura	# Records	342	926	763	540	325	481	424	0
	% of Ag Records	0.42	1.14	0.88	0.66	0.38	0.55	0.50	0.00
Yolo	# Records	1	21	0	0	0	0	0	0
	% of Ag Records	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00
Yuba	# Records	4	6	0	0	0	0	2	0
	% of Ag Records	0.07	0.10	0.00	0.00	0.00	0.00	0.03	0.00

NA: refers the unavailability of the data.

Appendix I

1. Field names in this document

MTRS	– Township, range and sections
Grower_id	- Grower identifications
Site_loc_id	- Site location identifications
Acre_planted	- values of the acres planted for the field and the crop
Acre_treated	- values of the acres treated for the field and the crop
Site_code	- Commodity code for which pesticide is applied on.

2. File naming conventions for the outputs in error checking

(a) Potential duplicated error - D_D90_45 refers to the duplicated error checking results for the data of 90 county 45.
(b) Potential acre planted error - A_D90_45 refers to the number of fields containing inconsistent values of acre planted for the data of 90 and county 45.
(c) Error records for the values of acre treated greater than acre planted - A_t_90_45 refers to the file containing the number of records that the values of acre treated are greater than the values of acre planted for the data of 90 and county 45.
(d) Not matched MTRS – nmd90_45 refers to the not matched number of MTRS and the records of not matched MTRS for the data of 90 and county 45.
(e) Missing location identifiers – M_m90_45 refers to the missing records for the MTRS for the data of 90 and county 45; G_M90_45 refers to the missing records for the grower_id for the data of 90 and county 45; M_S90_45 refers to the missing records for the site_loc_id for the data of 90 and county 45.
(f) Potential errors in MTRS - M_D9045 refers to the file containing the records that have the potential errors in MTRS for the data of 90 and county 45.
(g) Potential errors in county code - C_D90_45 refers to the file containing records that have the potential errors in county code for the data of 90 and county 45.
(h) Simple statistics – stati_45 refers to the simple statistics of the data.

3. Program descriptions

(see next page for the full description)

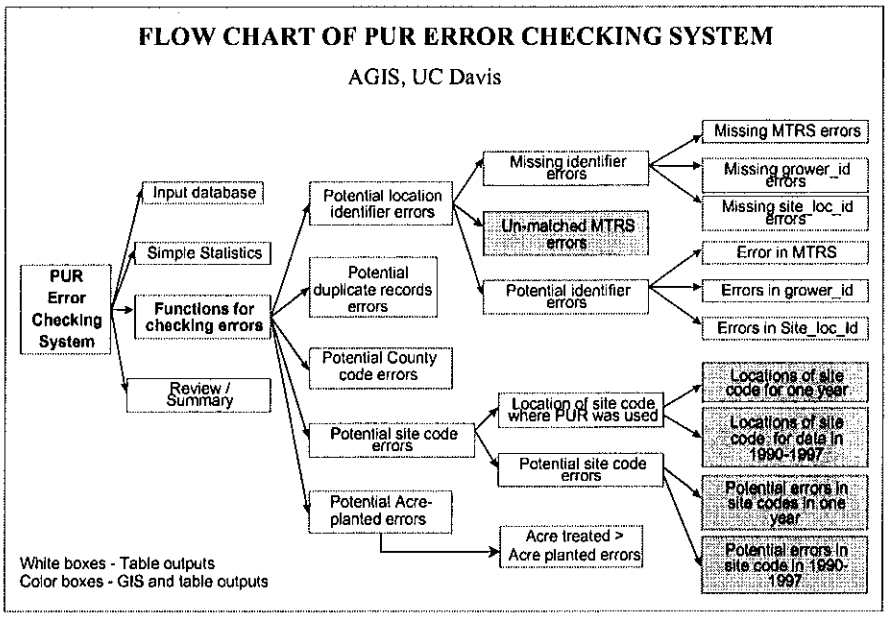
Descriptions of the Error Checking System for Spatial Attributes

Limei Yan, Yinyan Guo and Minghua Zhang
AGIS laboratory, UC Davis, September 15, 2000

With increasing awareness of the potential environmental impacts of pesticides, the database of Pesticide Use Records (PUR) is becoming more valuable to researchers, regulators, farmers and policy makers. To accurately assess the impacts of pesticides on the environments, the quality of the pesticide use data is extremely important. Therefore, we attempted to develop the computer system to check the errors relating to the spatial attributes.

This PUR error checking system is to examine pesticide use data of 1990 to 1997 and to identify the potential errors of various types that exist in the PUR database. Six functions were included in the system such as checking for duplicated errors, acre planted errors, commodity code errors, location identifier errors and county code errors. The location identifier errors included the errors in grower_id, site_loc_id and the township, range and sections. Diagram 1 illustrates the detail and relational structures of the system.

Diagram 1. Frame of PUR Error Checking System (ECS)



System Requirements and installation of PUR ECS

PUR ECS operates in the environment of Windows 95/98 or Windows NT (4.0 or later). The program was developed through standard Windows techniques, and requires software of MS Access 97 and ArcView3.1.

To install the PUR ECS, follow the steps listed below:

Organize the PUR data and get ready for error checking (data on CDs, that are organized by county code as directories and in the dBase (dbf) format; each county/directory contains multiple years of data for the same county).
Copy the county GIS coverage (shape file of PLSS – public land survey system, and DWR land use data file) from the specified CD to your local hard disk.
Create your working directory where you want to save your error checking results
Copy the files of PUR_ERROR_CHECKING_SYSTEM.exe and PUR_ErrorCheckingSystem90_99.mdb from the specified CD to your local working directory.
The setup process is completed and the program is ready to run in your local computer.

Getting Started with PUR Error Checking System

Select the correct path to each of the software location and prepare to run the program: double click "PUR_ERROR_CHECKING_SYSTEM.exe" from the Windows Explorer.
Following the instructions in the popup screen to locate the correct path for the Access.Exe program.
Following the instructions in the popup screen to locate the correct path for the ArcView.Exe program.
Locate the directory where county GIS coverages and the land use data from the Department of Water Resources (if any) are stored.
Specify the directory where the data are stored (use the CD including county 01 to verify the correct settings for the data, then insert the CD that contains the data for the county that you are interested to run the errors for).
Specify the working directory where you will save the results while running the program.

Screen of step 1.

AGIS

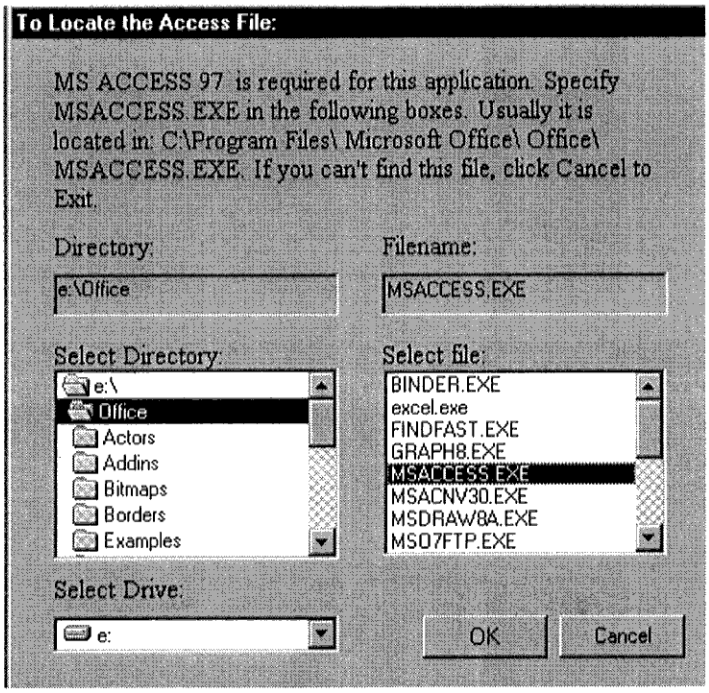
Pesticide Use Reports (PUR)

Error Checking System

Limei Yan *Yinyan Guo*

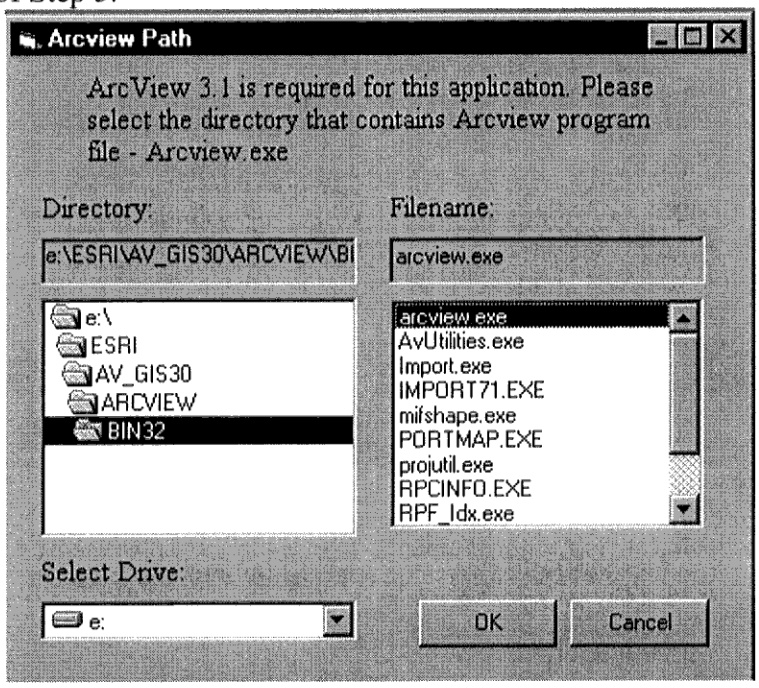
Agricultural Geographic Information System Laboratory
Department of Land, Air & Water Resources
University of California Davis

Screen of step 2.



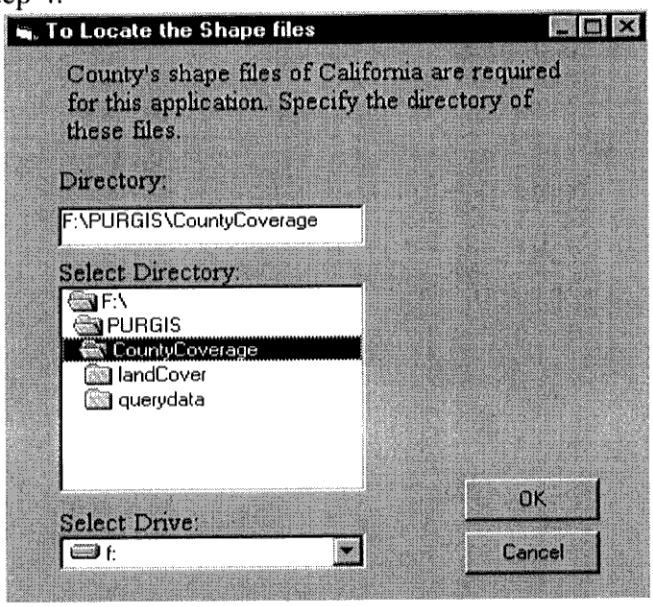
Use the Combo-boxes to select the location (drive and directory) of Access.exe program file. Find the Directory and program file for Access.exe. Then Double click the Access.exe program file. Or highlight Access.exe file, then click OK button.

Screen of Step 3.



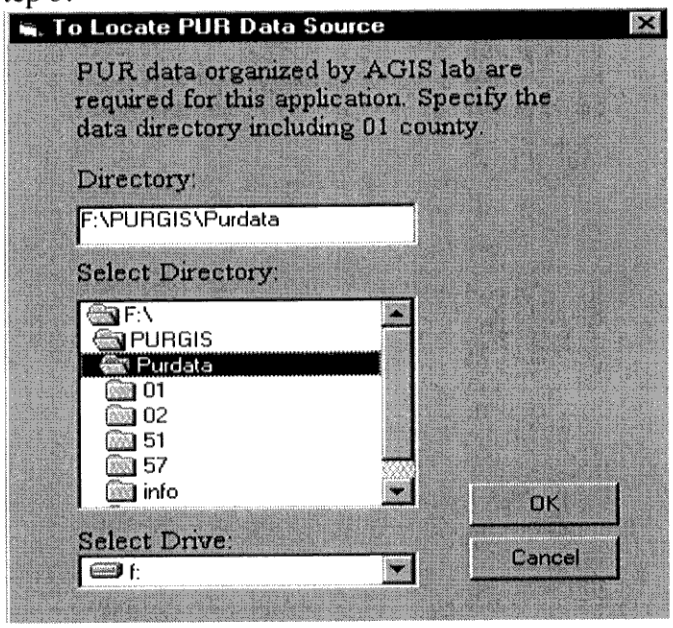
Locate the directory and the program file for ArcView.exe. Then Double click the Access.exe program file. Or highlight Access.exe file, then click OK button.

Screen of Step 4.



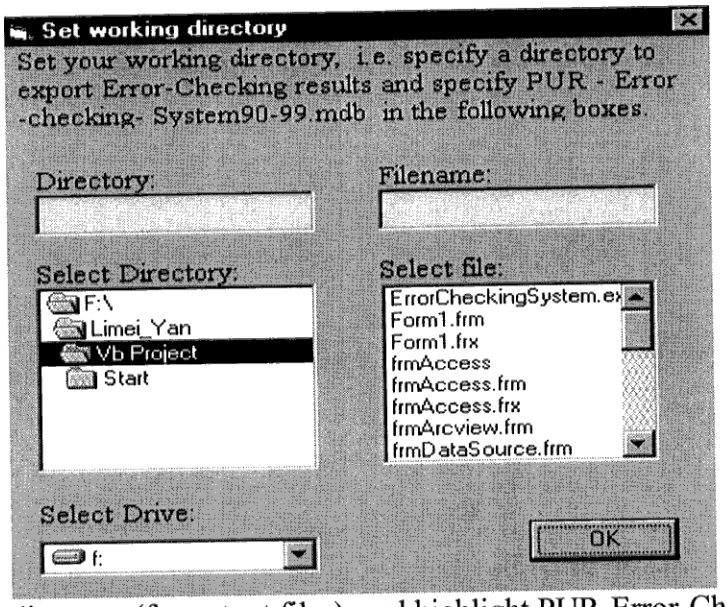
After find your County Coverage directory, click OK button.

Screen of Step 5.



Locate the data where the program will use.

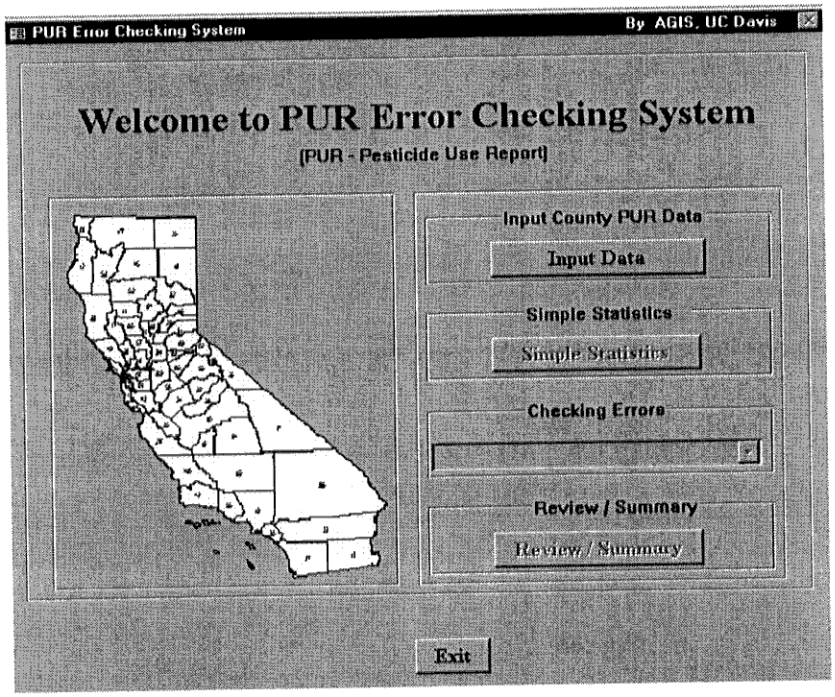
Screen of Step 6.



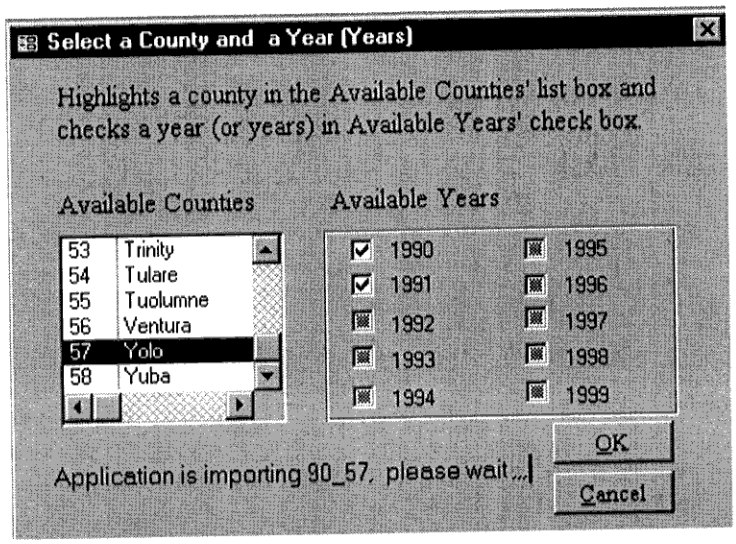
Select a working directory (for output files), and highlight PUR-Error-Checking System90_99.mdb.

Select County and Year(s) of data for inputting the data and calculating the simple statistics

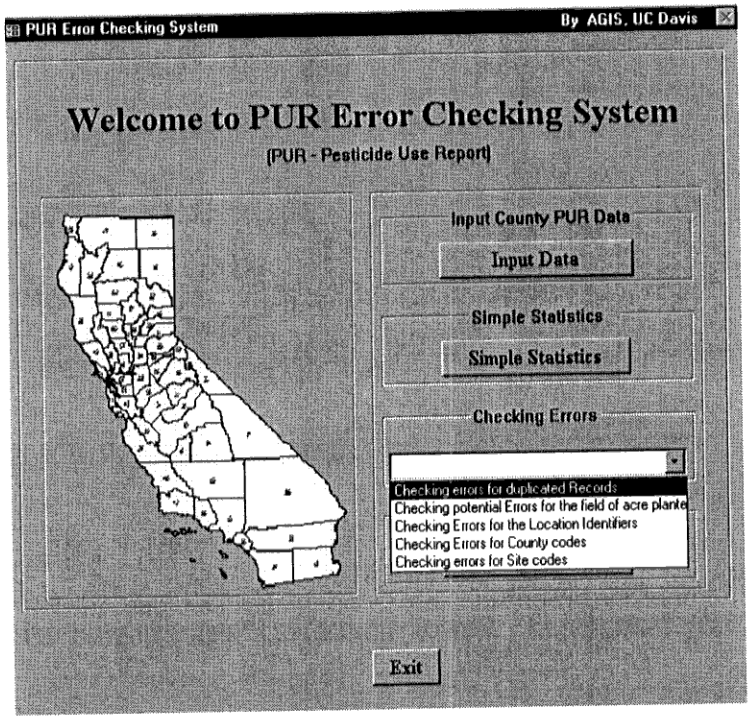
After all the necessary information is organized, the application will start running MS Access. One usually starts with inputting data to the system. You can select the county and year(s) of data you wish to check for the errors by highlighting or checking the boxes, respectively.



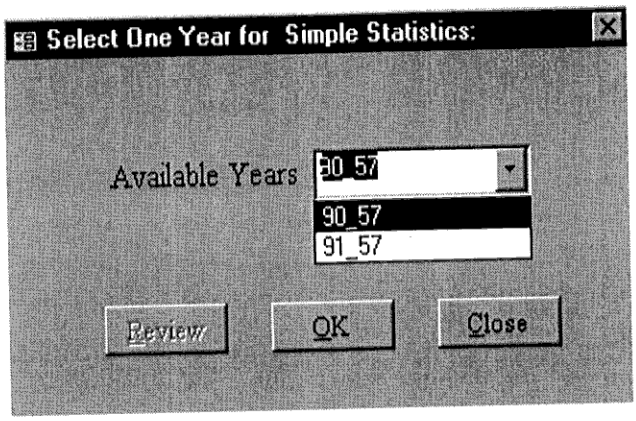
Press the button of "Input Data" in the above screen to launch the following form for the selection of county and year(s) of data.



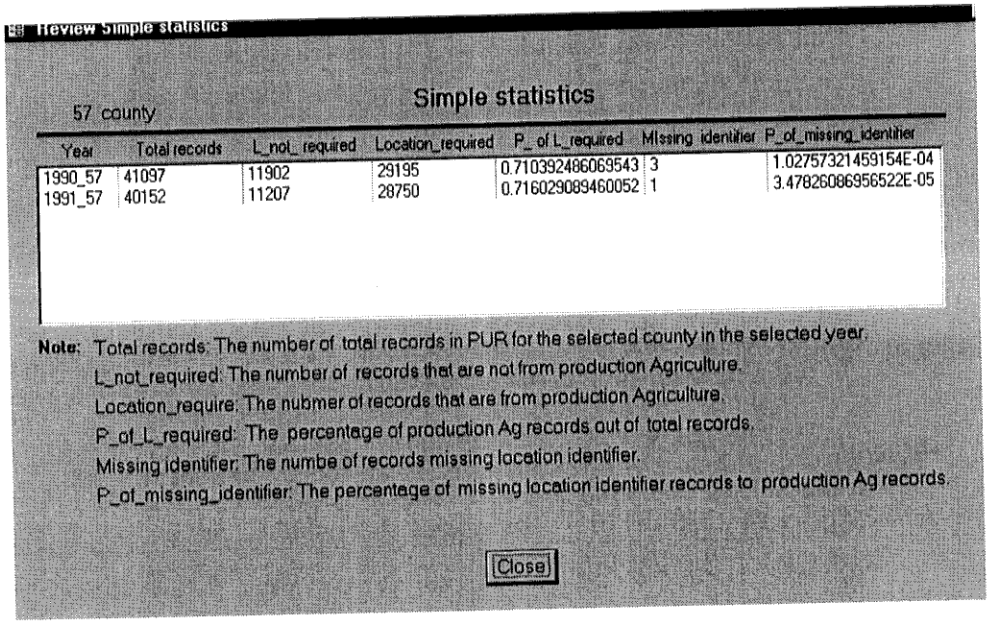
After inputting the data, one can press the button of "Simple Statistics" in the following screen to obtain the summary for the total records, agricultural production record, percent of the agricultural records etc. simple statistics. These results are saved in the file under your working directory that you setup previously.



As soon as the button of "Simple Statistics" is pressed, the following form will appear on screen to request a selection of a year for the simple statistics.



Then users can select the year and press the "OK" button. Repeat the same process for all the years that users want to check for, then press the button of "Review" to view the results. The following screen provided a view of the simple statistics for two years of data in Yolo County. The explanation for the columns is described in the note of the form below.



Basic PUR Error Checking System Functions

We designed the system to check for the errors of the following aspects:
Duplicate records, i.e., any records containing the same information for the following nine fields, then we extract the duplicate records and save them into a file for further examination. These nine fields include grower_id, site_loc_id, acre_planted, acre_treated, prodno, chem-code, lbs_chm_us, applic_dt, and site_code.

Records with potential errors in the values of acre_planted. First of all, we select the records that were from agricultural productions. Then we extract the records that have different values for the acre_planted for the same combination of grower_id, site_loc_id

and site_code. The rational is that the same field growing the same crop during the same season should have the same values for acre_planted. Therefore, if there are different values for the field of acre_planted in the combination of grower_id, site_loc_id and site_code, then it has the potential that one or more than one records contains wrong information in acre_planted. If the values of the field for acre_treated are greater than the values of the field for acre_planted for the same field designated by grower_id, site_loc_id and MTRS, then we concluded that the records with larger acre_treated values are likely the fields with reporting mistakes.

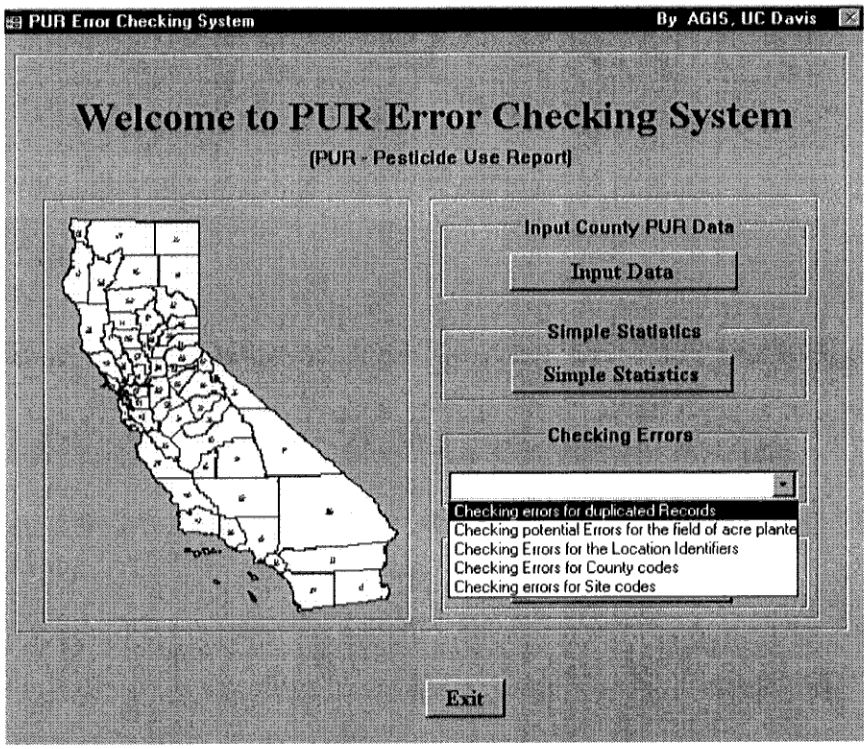
Location identifier, this item includes not-matched MTRS, missing location identifiers and mistakes within identifiers.
Not matched means that some of the township, range and sections in the PUR does not belong to the township, range and section for the county. We compared the existing MTRS in the PUR with the county GIS MTRS and then extracted the not-matched MTRS records and saved these records in a file in the working directory for further examination.
Missing location identifier refers to the missing information in the PUR including missing grower_id, or site_loc_id or MTRS in any of the records. We also saved these records for further checking in determining exact errors.
Mistakes within the identifiers refer to the potential mistakes in each of the location identifiers. These include grower_id, site_loc_id and MTRS. We attempted to find the records that contained different grower_ids for the same combination of site_loc_id and MTRS, different site_loc_ids for the same combination of grower_id and MTRS as well as different MTRS for the same combination of grower_id and site_loc_id. All these records are stored in the file for further examination.

Records with inconsistent county code in the first two digits of the grower_id field. For example, Yolo County should have a code of 57, the program will check for any records that are not 57 in the first two digits of the grower_id field, and then save these records into a file for further examination.

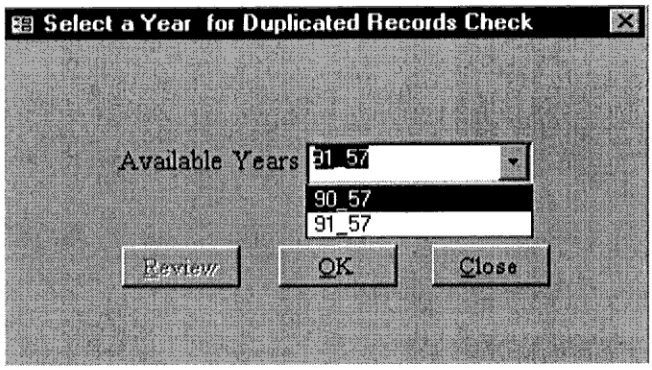
Site Code potential errors.
The site code potential errors refer to the possible mis-report on commodity use for some of the pesticides. The program allows users to select the commodity code and make maps to check for the potential errors on site code. The potential errors may not be obvious to these people who are not familiar with the area. However, it will be clear to the people who have a good knowledge about the area or region.
If the landuse from the Department of Water Resources is available for the county of interest, the program allows users to select the site code (commodity code) to compare with the landuse map for the discrepancies on the actual physical locations. The differences of the spatial locations will be mapped after the comparison. The difference map may mean that the report of the site code in PUR contains potential errors in these locations, or it could mean the landuse map is out of date. Therefore, the maps are only references in assisting the determination of whether there are errors in these locations for the commodity of interest.

Detail Screen Displays for Each Functions

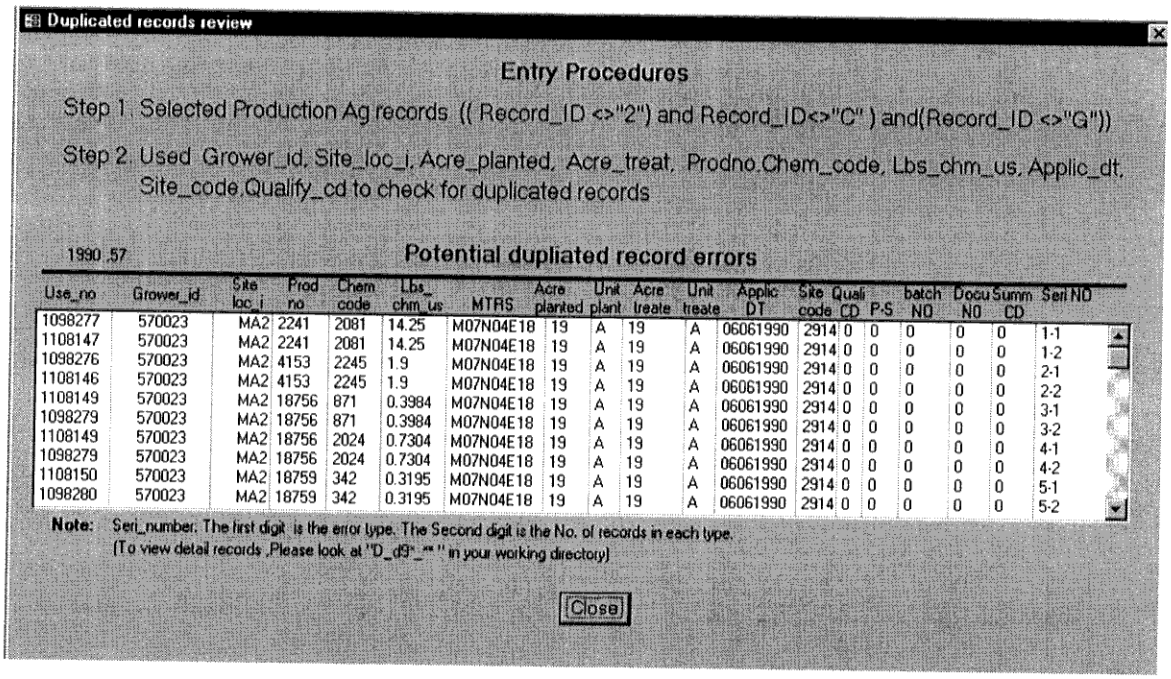
1. Checking for duplicate records:
Highlight the duplicate records in the Checking Error box.



Enter the year of interest:

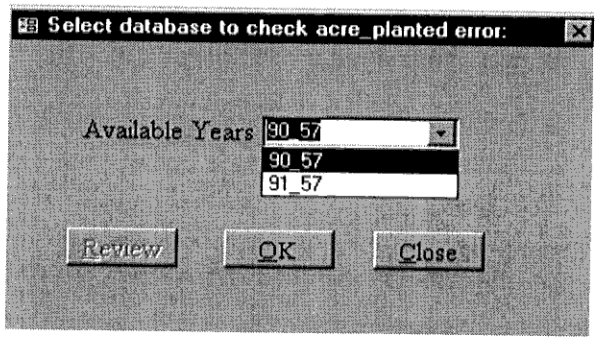


Review the results after checking the records for each year. Or repeat the process until all the data from multiple years are completed and then press the "Review" button to view the results for multiple years. One can check the files generated during this process for each year under the working directory.

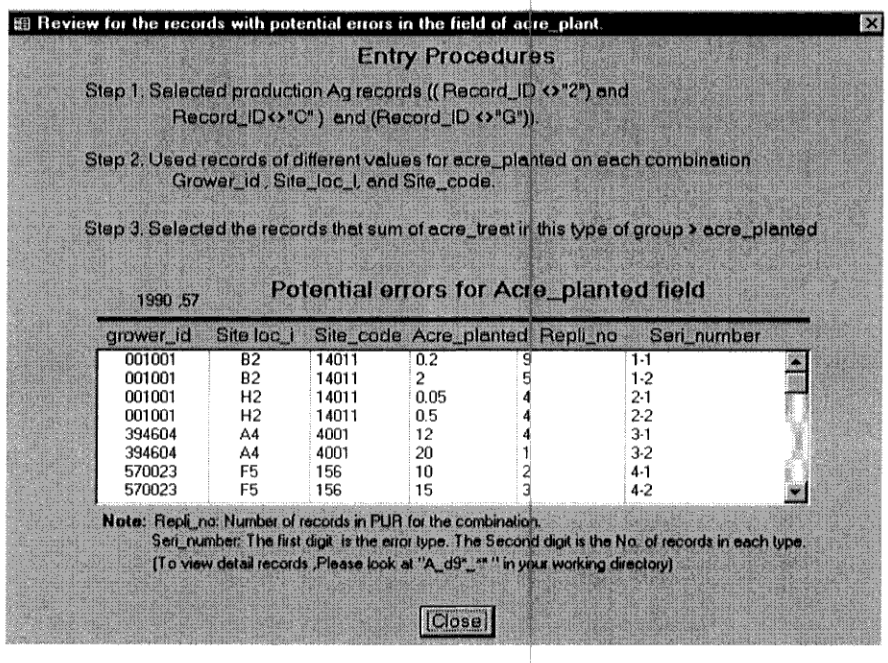


2. Checking for potential errors in acre planted:

If a user selects "Checking potential errors for the field of acre planted" in the comb box, the following form will appear for you to select the year of the data.

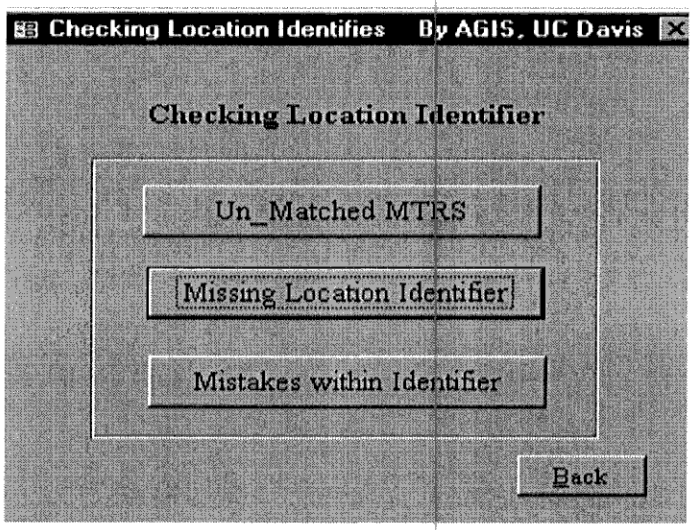


After a user selects the year, the application will check the potential acre planted error for the selected year. "Review" button then will be activated. If a user presses the "Review" button, a form similar to the following will appear to show the results.

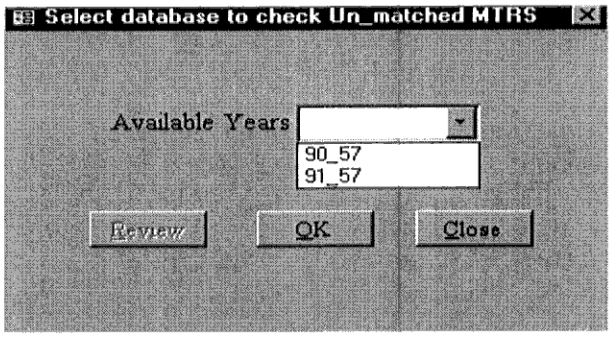


3. Checking errors for the location identifiers.

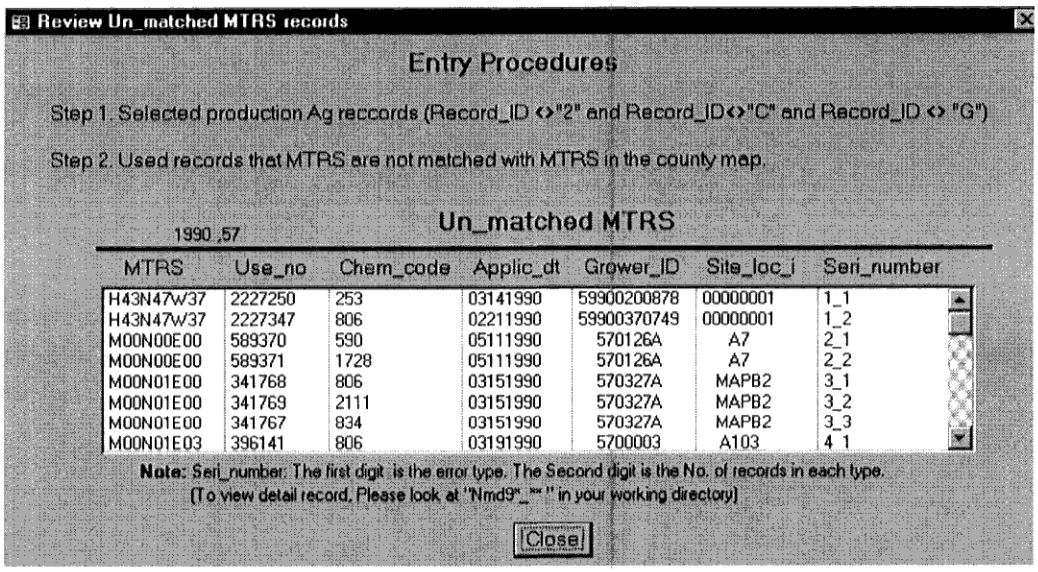
The screen below displays the submenu for checking for the errors in location identifiers.



Un-Matched MTRS (which, called as "extra MTRS", existed in the PUR database, but not present in county GIS coverage). By pressing this button, the application will first input the corresponding county GIS coverage attribute file in DBF format. Then one form with combo box will pop up for the selection of a preferred year to check for errors in this item:

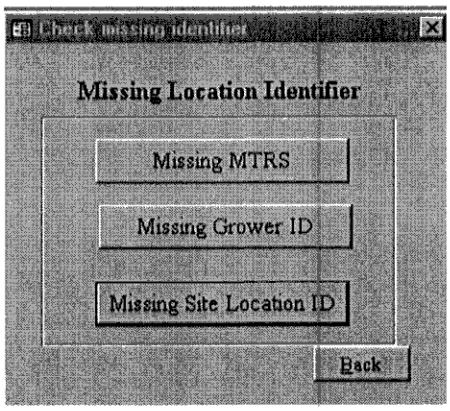


After a user selects the year, the application will find un-matched MTRS comparing with county GIS coverage. If some Un_matched records are found, the "Review" button will be activated. By pressing this "Review" button, a user can view the records of the un-matched MTRS.

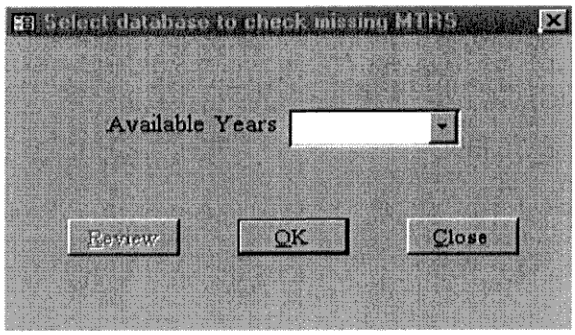


If one found no un-matched records, a message box will be displayed on the screen indicating "there is no un-matched records in the database".

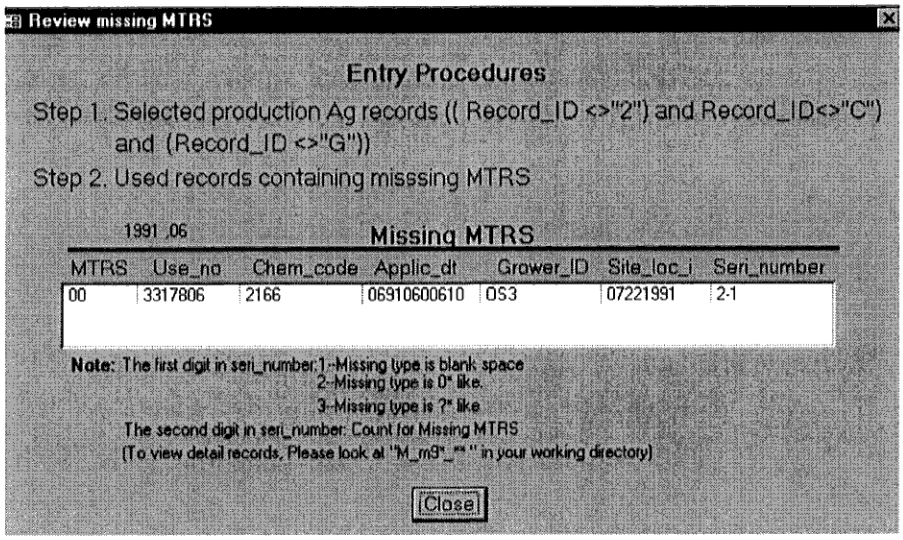
Missing Location Identifier (Missing MTRS or grower id or site location id).



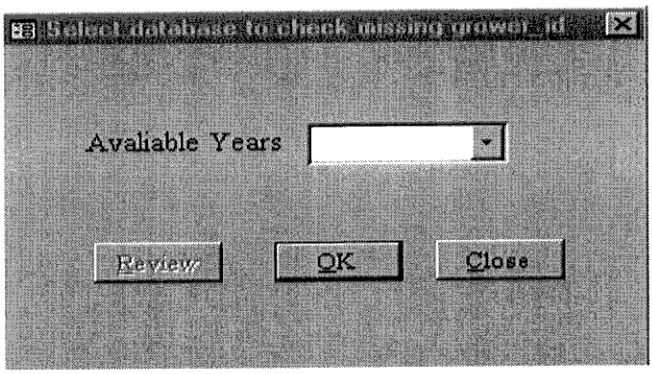
For checking the missing MTRS, pressing this button, the application will prompt to select the year. Once the year is selected, the program will find the missing MTRS in the selected database.



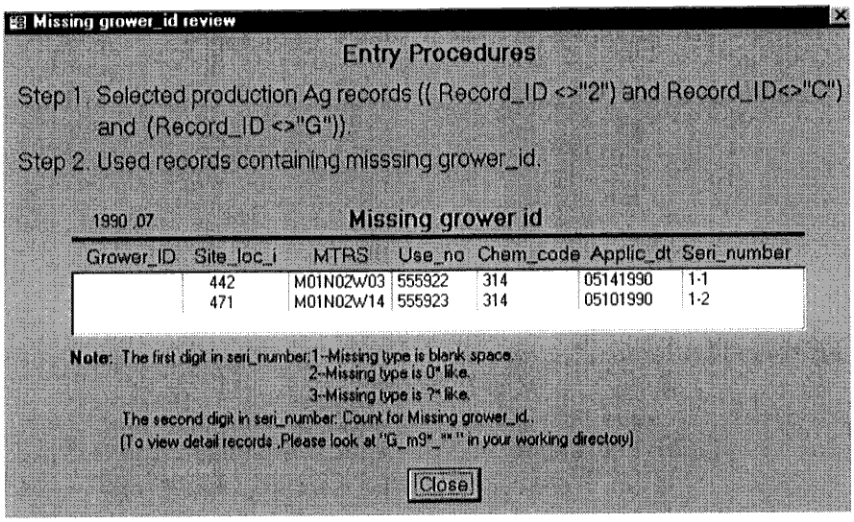
If the program found missing MTRS among the data of the selected year, the "Review" button will be activated. By pressing the "Review" button, a user can view the Missing MTRS records.



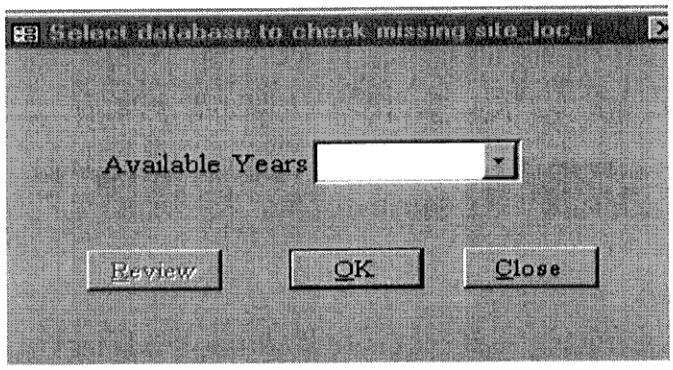
Checking missing grower_id: Pressing the button of "Missing Grower_id", the application will prompt a user to select the year for the database. Once a user enters the year, the program will find the missing grower id in the selected database.



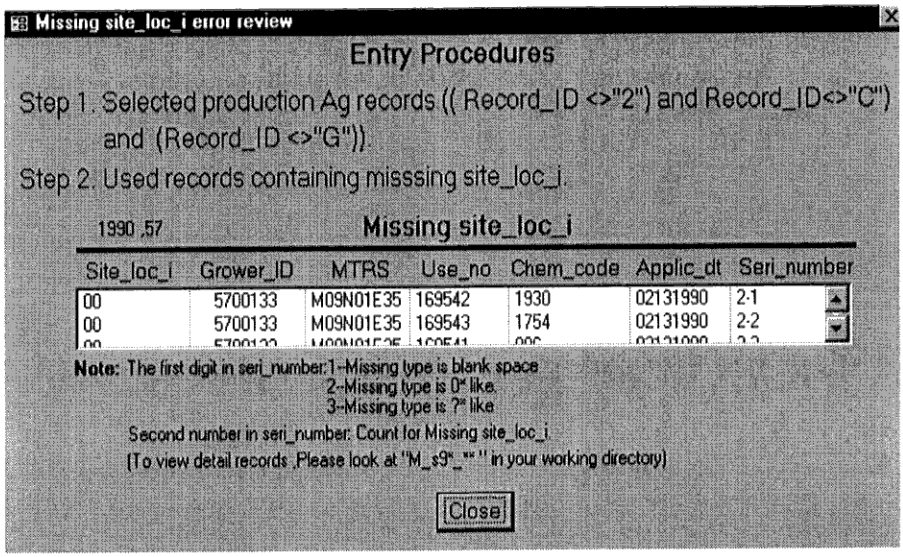
If there are records with missing grower id in the selected year, the "Review" button will be activated. By pressing the "Review" button, users can view the "Missing grower id" records.



Checking missing Site Location ID: Pressing this button of "Missing site_loc_id", the application will prompt a user to select the year of the database. Once a user enters the year, the program will find the missing site location id in the selected database.



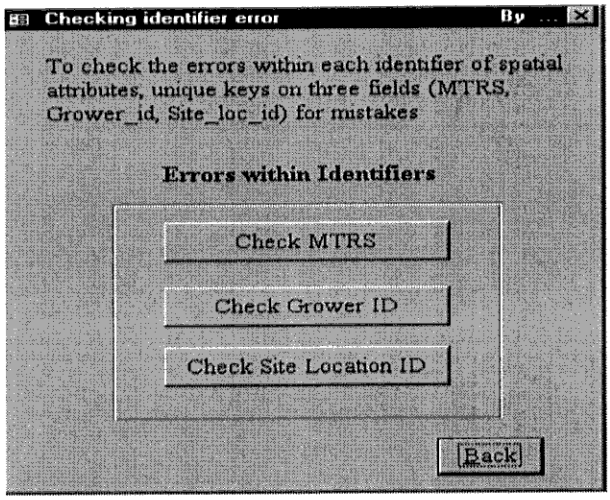
If there are missing Site_loc_i in the selected year, the "Review" button will be activated. By pressing the "Review" button, users can view the Missing Site_loc_i records.



Use "Back" button to return to the "Check Location Identifier" form.

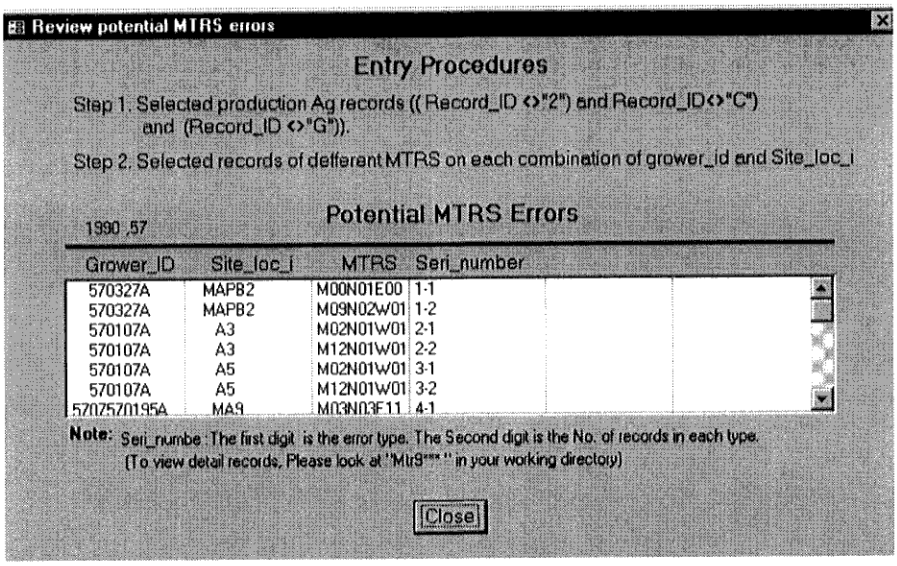
(3) Mistakes within Identifiers

This form will be displayed after selecting the button of "Mistakes within identifiers".

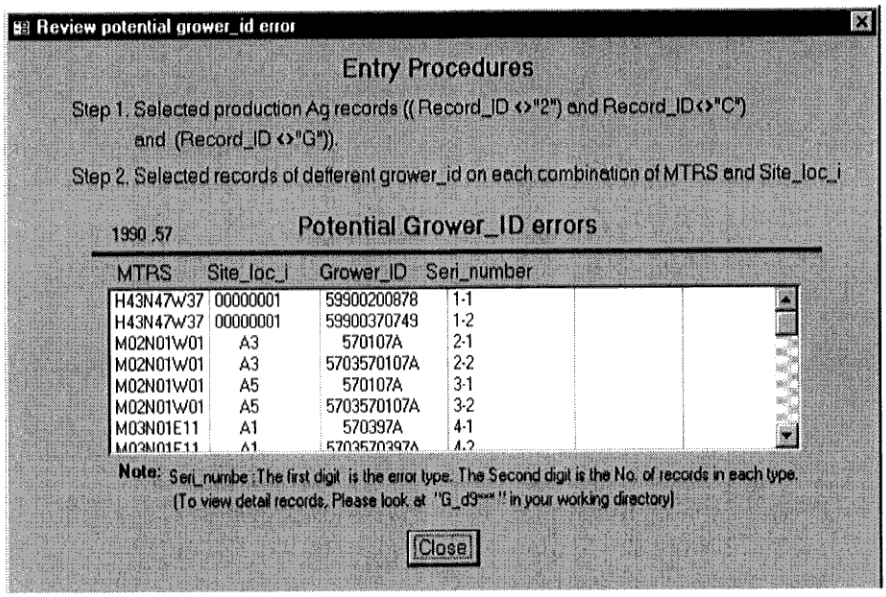


Users can press "Check MTRS" button to obtain the potential MTRS error records and press "Check Grower_id" button to obtain the potential Grower_id error records, or press "Check Site location ID" button to obtain the potential site_loc_i error records. After one finishes checking any of these functions, the "Review" button will be activated, and one of following forms can be viewed by pressing "Review" button.

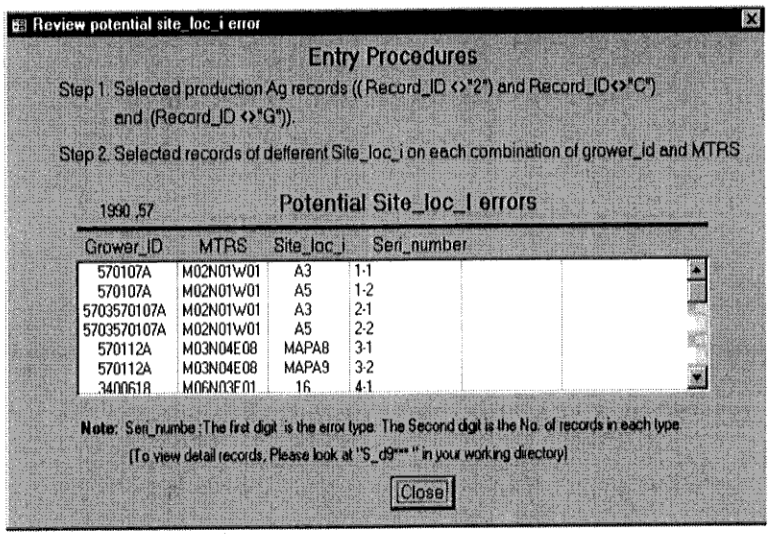
Review Potential MTRS Errors:



Review Potential Grower ID Errors:

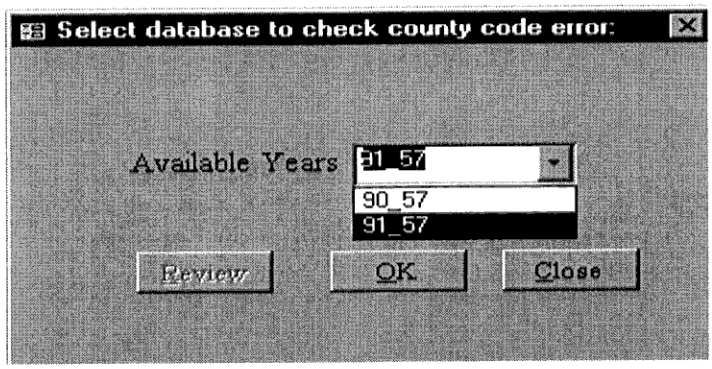


Review Potential Site Location ID error:

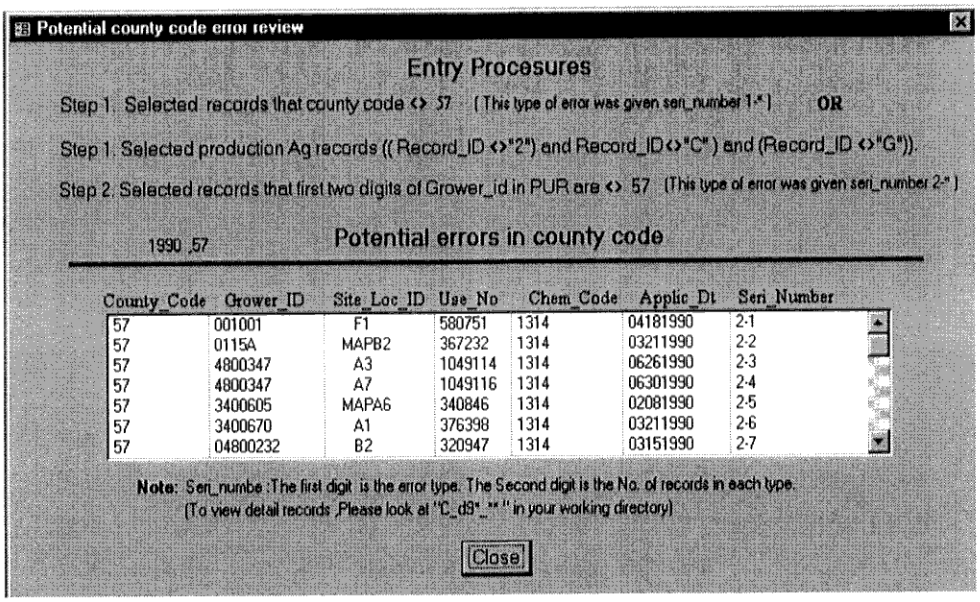


4. Checking errors for County codes:

After selecting the function for checking county code, users need to select the year to check for the errors in the data.

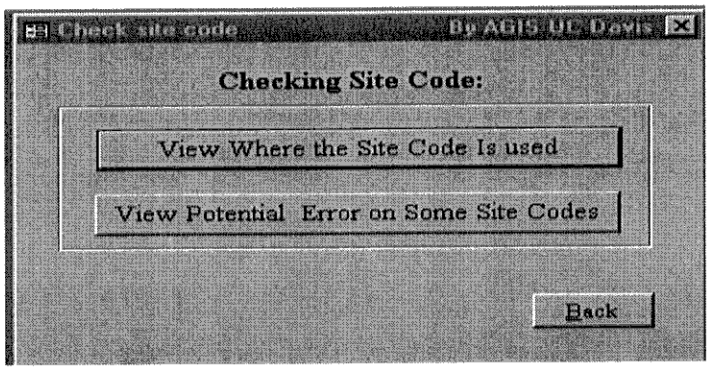


After a user selects the year, the application will check for the potential county code error. If there are county code errors, the "Review" button will be activated. By pressing the "Review" button, the following form will be displayed on screen.

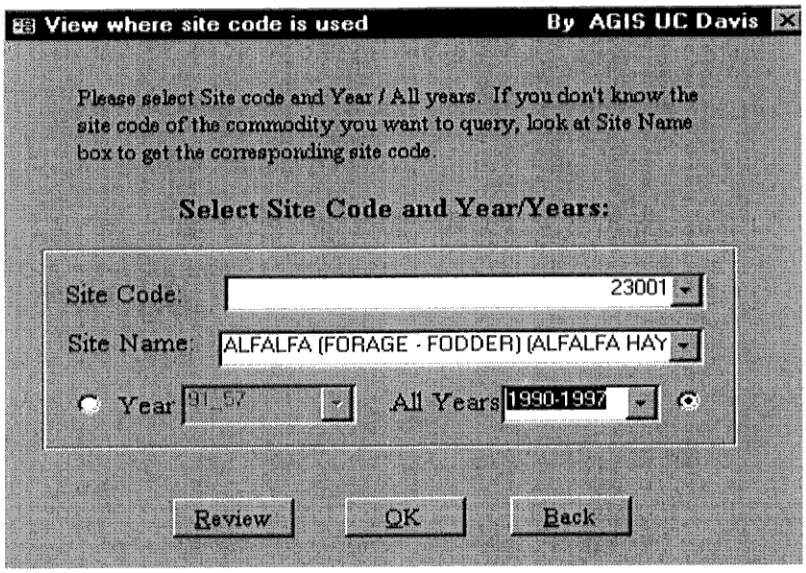


Checking errors for Site codes:

In the comb box, users need to select one of the two choices.

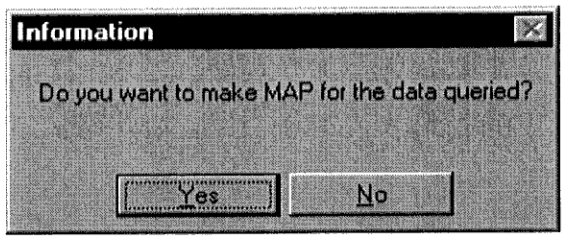


If a user pressing the "View Where the Site Code Is Used" button, the following form will be displayed on screen to request an input for site code. The site name is informational to assist users to locate the correct and corresponding site code for the commodity name of their interest. However, site name cannot be used to link the data. Therefore, users have to select the correct site code for checking the errors in site code.

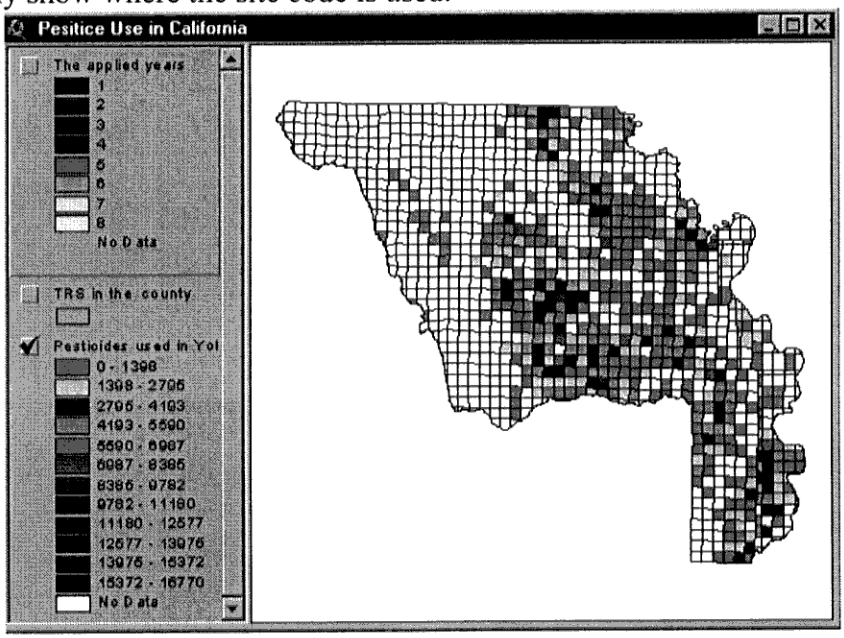


.

If a user wants to view where site code is used in 8 years (1990-1997), one should select the site code comb box and all years comb box. Press the "OK" button, the application will query for these records. Then the application will ask whether making a map in a message box.



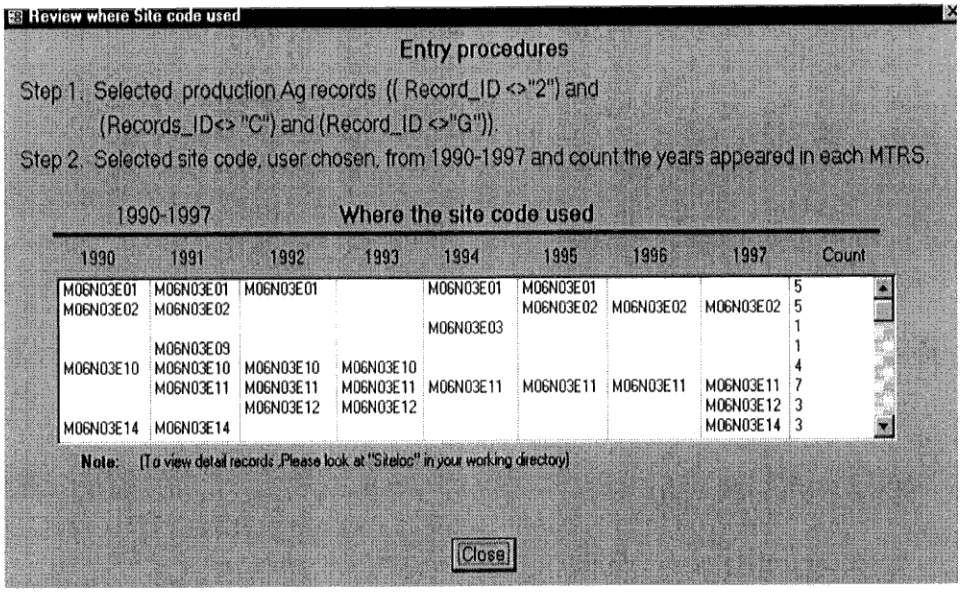
By pressing Yes, the application will automatically start the ArcView.exe, and automatically show where the site code is used.



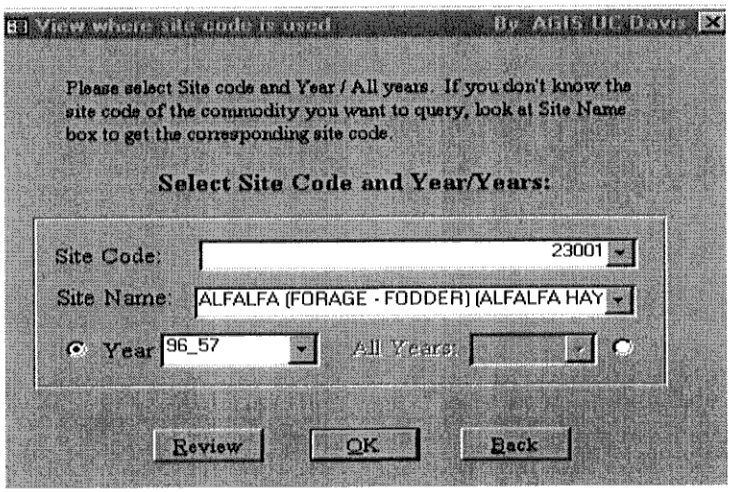
The above graph is an example from the pesticide use for alfalfa fields in 1990 to 1997 in Yolo County.

Pressing "Exit" button in ArcView in the View environment to exit the ArcView, or saving the map in the layout before exiting ArcView.

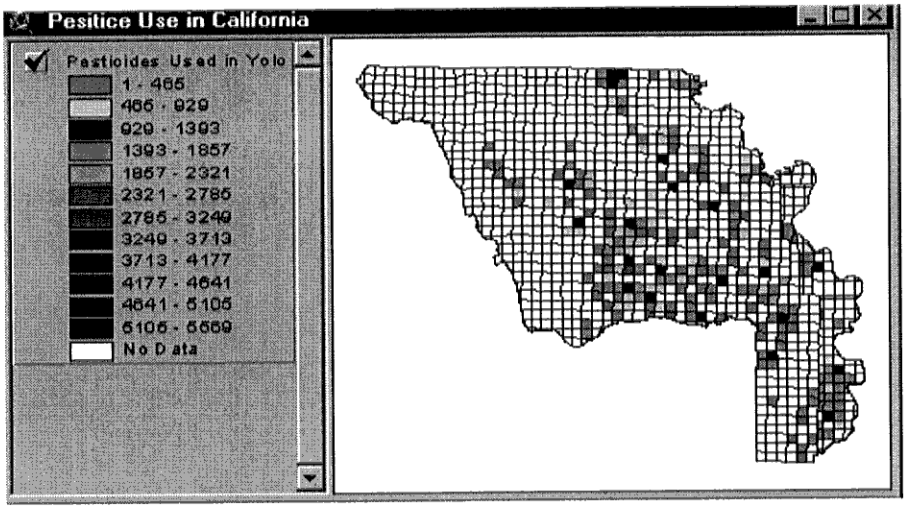
The "Review" button in "view where site code is used" form will be activated. By pressing the "Review" button, the following form will show on screen to inform where site code, in terms of township, range and section, is used in each year from 1990-1997.



If a user selects a year in the comb box as indicated in the following form, only one year's data will be queried.

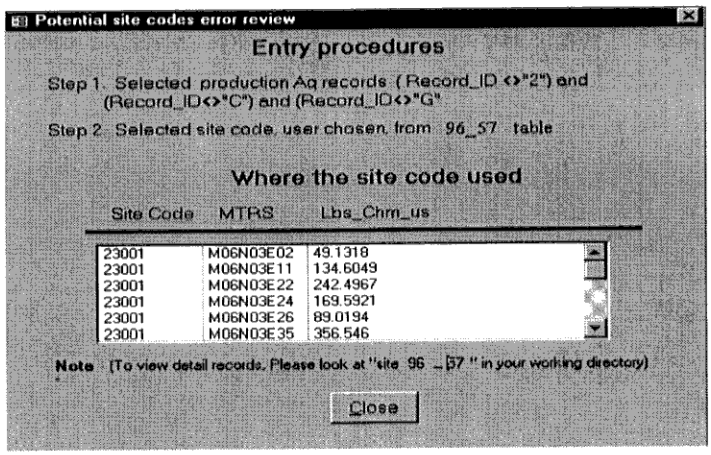


If a user wishes to view the map, the following map will be shown on screen in ArcView.



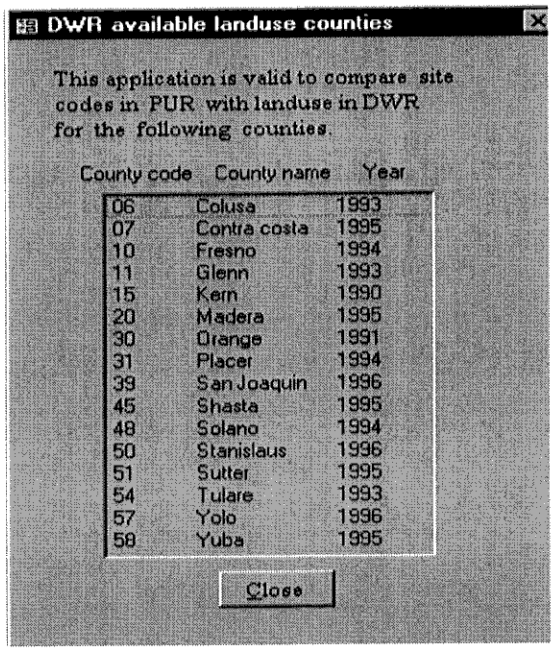
This graph displayed the pesticide uses in 1996 on alfalfa fields in Yolo County.

Press "exit" button to exit ArcView program. Then press the "Review" button to review the results.

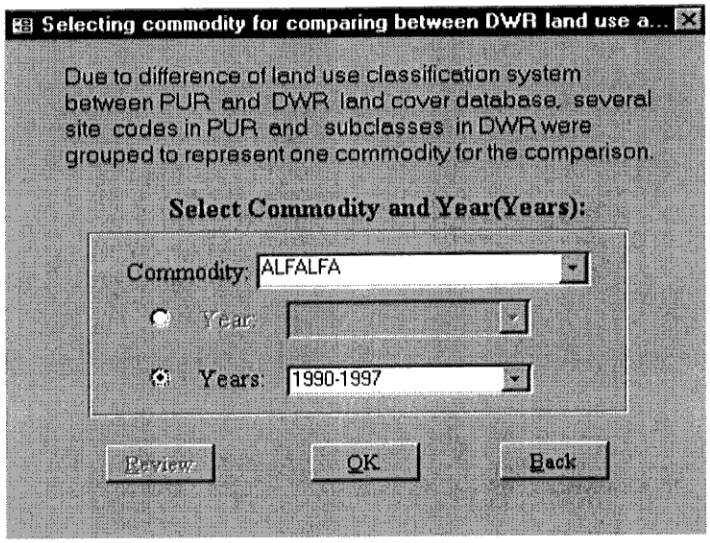


Comparing the site code with the landuse from Department of Water Resources: Assuming the landuse from DWR is more accurately reflecting the actual field boundaries, we can compare the site code in the PUR database with the landuse from DWR to see whether site code is correctly reported in the PUR. As noted previously, there may be time differences in landuse map and PUR data. The differences between the two databases only serve as a reference.

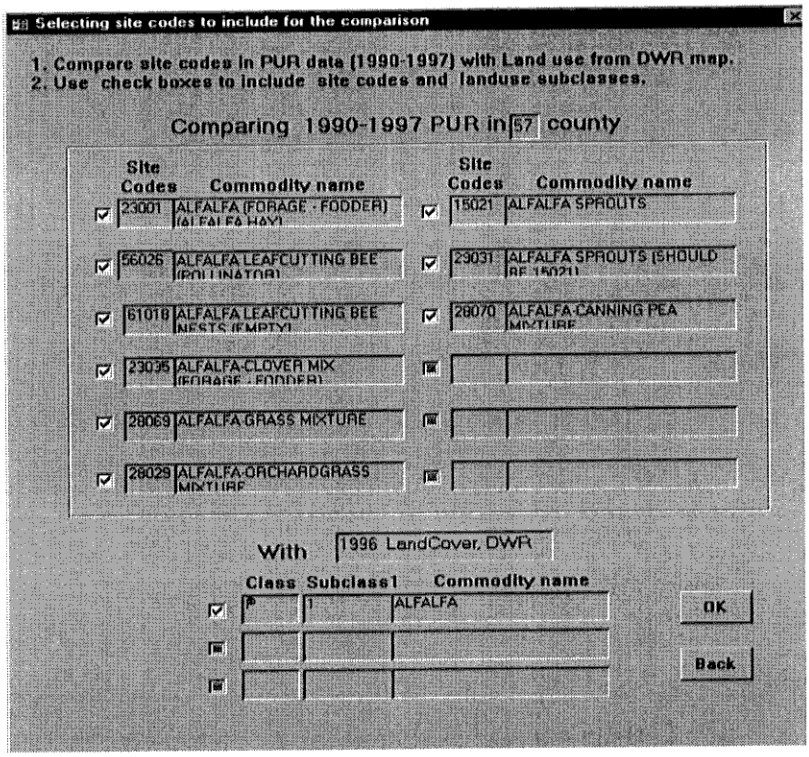
The following steps allow us to check for the site code errors in the PUR when comparing the site code with the landuse from DWR. However, only 16 counties of landuse from DWR are available to use at the time. Therefore, if a user presses "view potential error on some site codes" button, when there is no landuse coverage available in this county, the following form will pop up to show you the available landuse data.



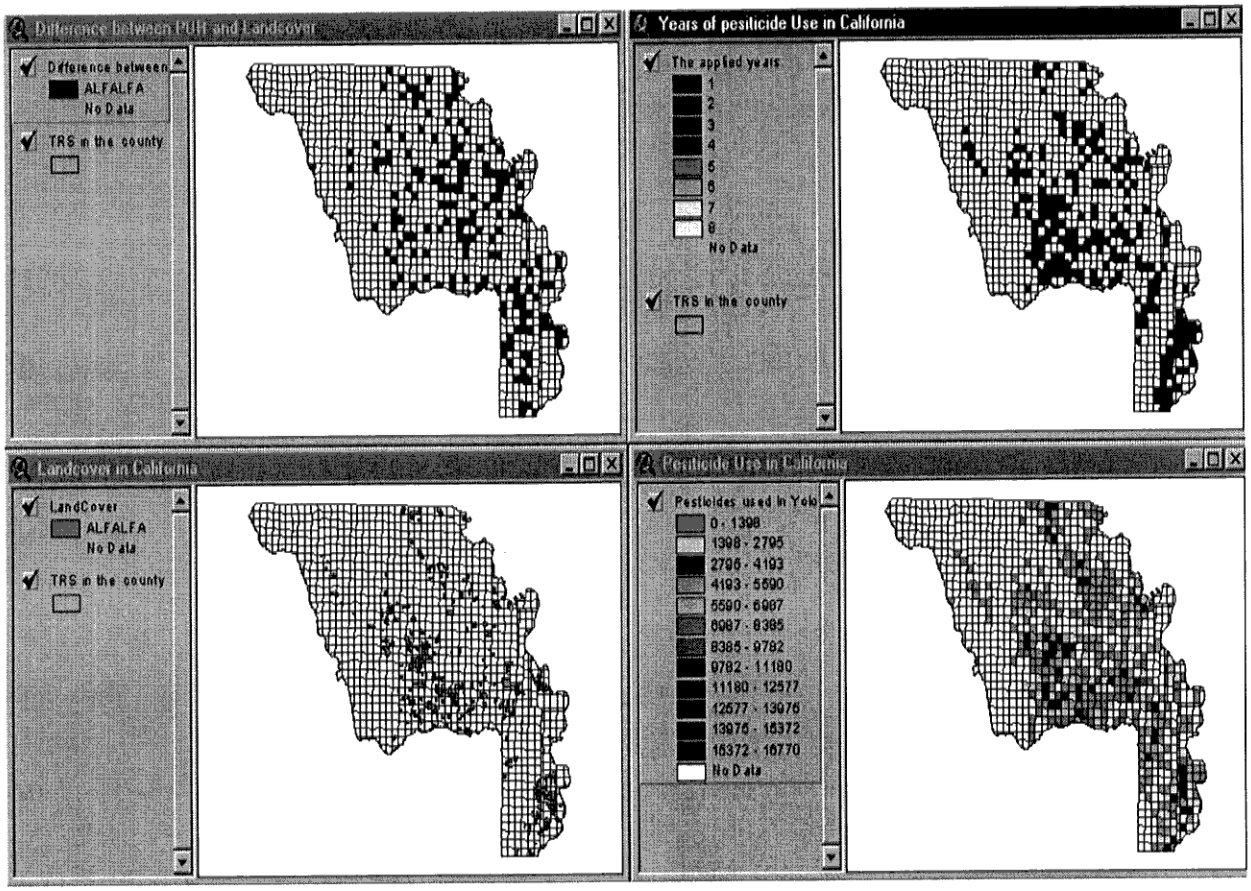
If a user presses "view potential error on some site codes" button and when there is a landuse coverage available for this county, the following form will be displayed.



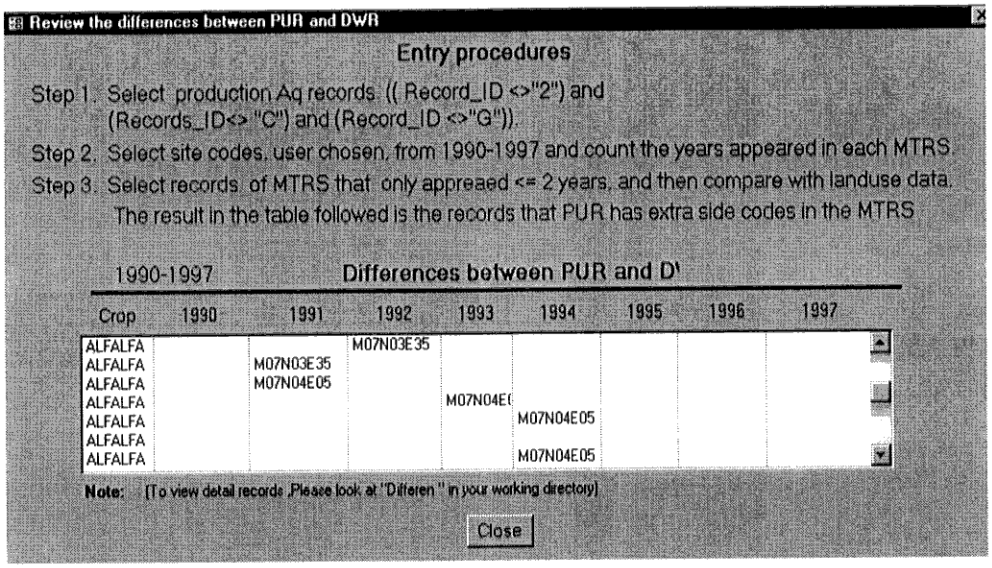
If one selects "years" in the comb box of "Commodity and Year", the following form will pop up to the screen.



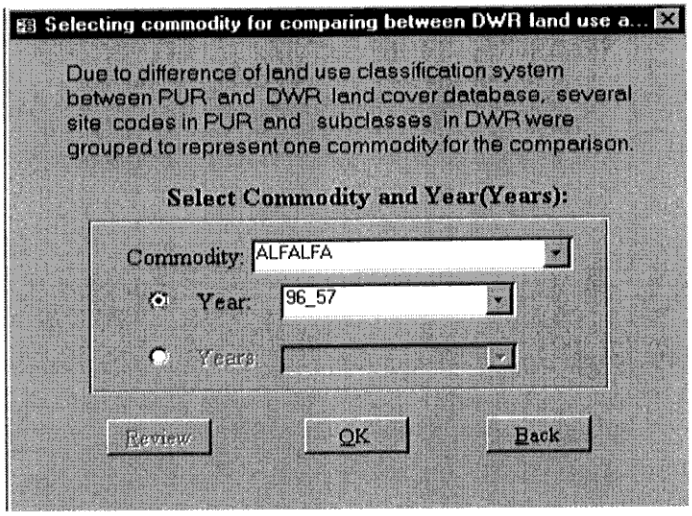
Following the instructions in the comb box, one can select site codes and click "OK" button, the application will compare PUR with DWR landuse coverage. If one selects "Yes" to display the maps, the following results will be shown in ArcView.



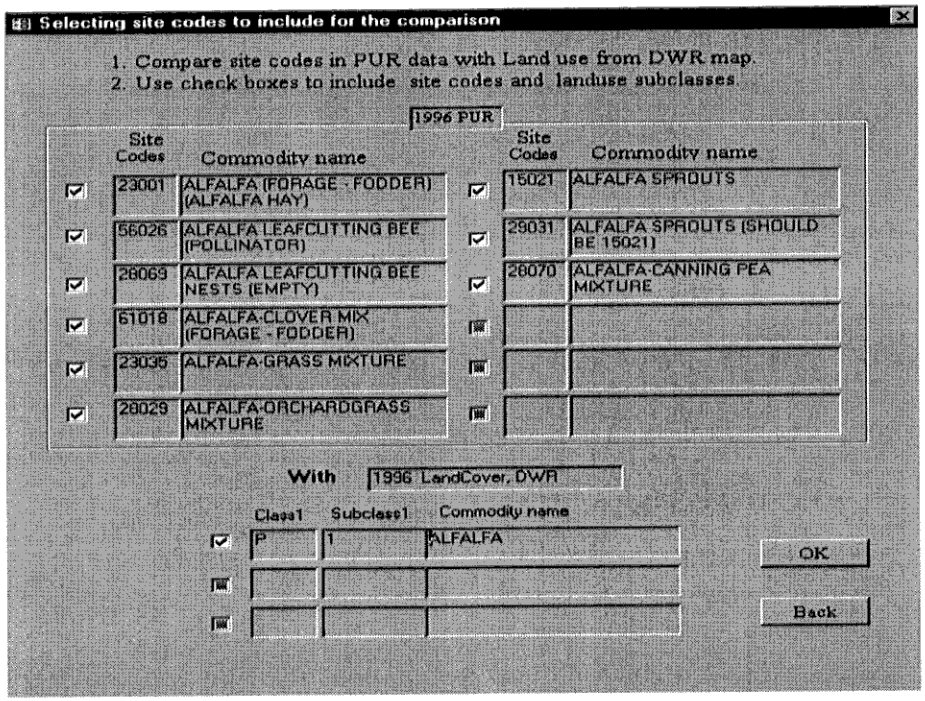
A user can get the difference between PUR and DWR by pressing the "Review" button in "Selecting commodity for comparing between DWR land cover and PUR site codes" form to get the following results.



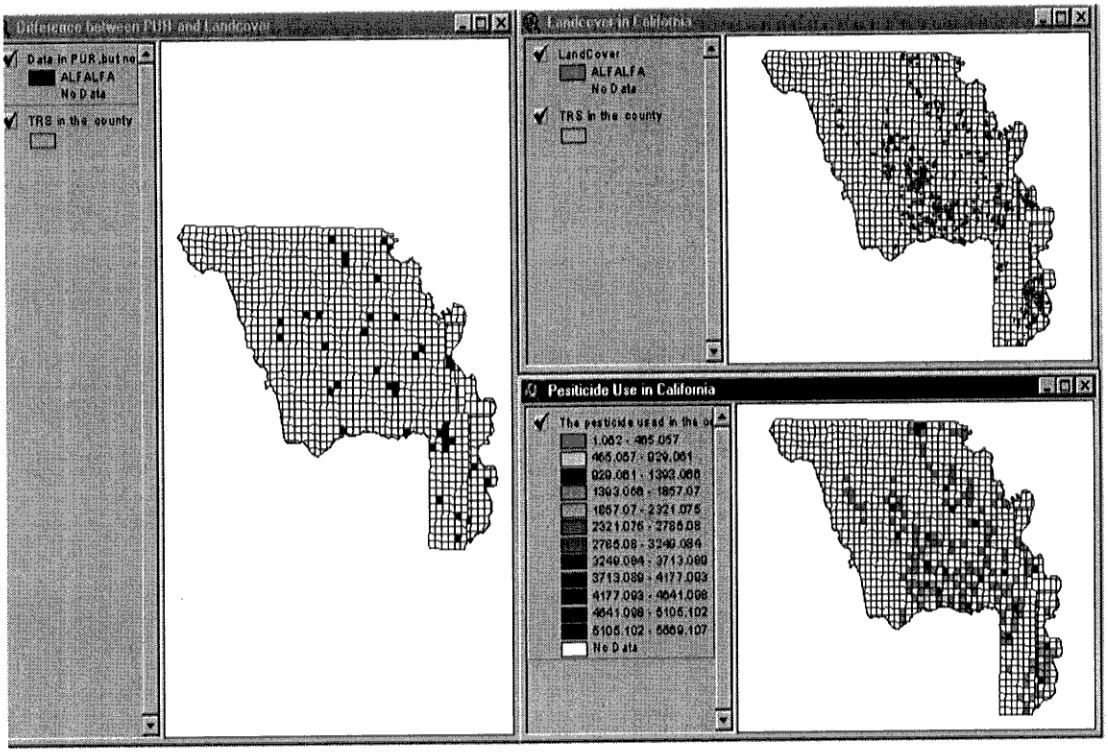
If a user selects "year" comb box and commodity comb box in the following form.



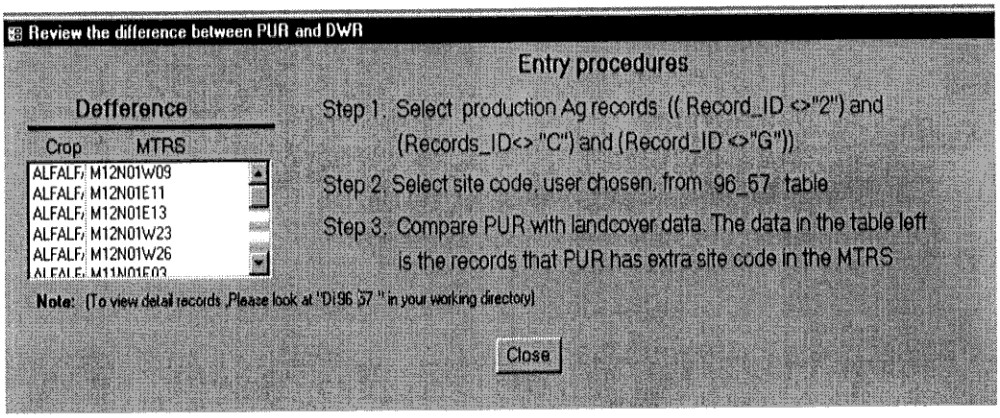
By pressing "OK", the application will pop up the following form.



By selecting site codes and click "OK" button, the application will compare PUR with DWR landuse coverage. If a user selects "Yes" do display maps, the following results will be shown in ArcView.

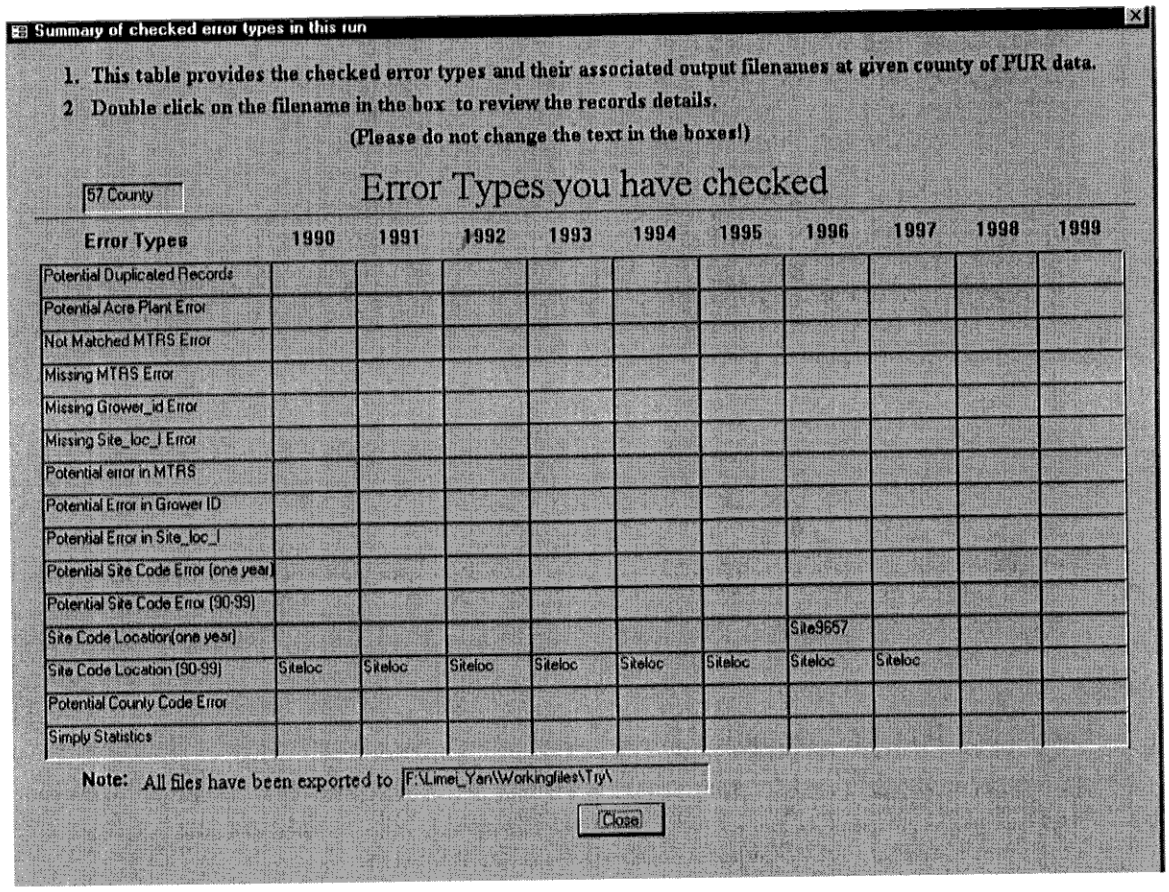


A user can get the difference between PUR and DWR by pressing the "Review" button in "Selecting commodity for comparing between DWR land cover and PUR site codes" form to get following results.



If all the MTRS in"View Potential Error on some site codes" section is not matched with the county GIS coverage MTRS, no map of potential errors will be displayed on screen, and the legends in the ArcView view is shown as " ".

In summary, a user can view the Review/Summary to get error checking results in this step by pressing "Review/Summary" button in main form. The following form will be displayed on the screen.



Note: Due to the size capability limitation in MS Access, if "Access has reached maximum size" error message appears, Please compact the database and rerun the desired checking functions. For the county with large PUR data records this application may have its limitation.

APPENDIX D

Pesticide Use Report Loading and Error Handling Processes

PESTICIDE USE REPORT LOADING AND ERROR HANDLING PROCESSES

Larry Wilhoit

March 29, 2001

Table of Contents

List of Tables and Figures

Introduction

The California Department of Pesticide Regulation's (DPR) Pesticide Use Report (PUR) is probably the largest and most complete database on pesticide use in the world. A system to collect information on California pesticide use has been in operation in some form for more than 40 years, but the current, full use reporting system began in 1990. The PUR database discussed in this report refers only to the data collected since 1990.

The PUR contains information on nearly all production agricultural pesticide use and on some non-agricultural use in California. The data collected includes what pesticide product was used, the date it was applied, which field was treated, and location to a square-mile section. Production agricultural use includes applications to growing crops, agricultural fields, and most applications to forest trees and ornamental turf. For brevity, these uses will all be referred to as "agricultural uses." Other pesticide uses reported to DPR include post-harvest commodity treatments, rights of ways, landscapes, structural use, and other non-agricultural uses by commercial applicators. These heterogeneous applications will be referred to as "non-agricultural use."

DPR expanded pesticide use collection in 1990 primarily to more accurately assess dietary risks from pesticide exposure. However, the PUR is now used for a wide variety of environmental and public health purposes, including risk assessments, promoting farm worker health and safety, analyzing human exposure patterns, protecting threatened and endangered species, monitoring and investigating environmental issues, and improving pest management. The PUR is used extensively by state and U.S. governments, universities, farmer organizations, the pesticide industry, and public interest groups.

This report describes the operation of only one part of this data collection system. It provides some of the technical details of the computer program and procedures for loading PUR data received from the counties into an Oracle database and checking the data for errors. This program is known as the loader program.

The PUR Process

To understand the function of the loader program, it will help to understand the larger PUR process that it is part of. Data collection starts with pesticide users filling out a form to report each pesticide application. Most of these forms are paper but some are electronic. These reports are sent to the county where staff enter the data into a local database. Every few weeks most counties produce a text file containing the latest data collected. This file is either emailed or copied to a floppy disk and mailed to the state headquarters at DPR. At DPR, the loader program runs every day and checks if any new county data files have been received. If the program finds a new file the data are run through a series of error checks. Error-free data are entered into a database table and errors are recorded in another table. The error table is used to generate an error report, listing each error. This is printed and sent back to the county where the data came from. Staff at the county must try to correct each error in this error report. This usually means

that they must find the original pesticide use report submitted by the pesticide user. The correction is made to the county database and marked on the printed error report, which is sent back to the DPR where staff make the correction to DPR database.

Generally the counties send the PUR data to DPR every few weeks, but sometimes it can take several months. By April, DPR should have nearly all the county data from the previous year. However, the error corrections take several more months. The final version of the PUR is typically available in December and contains the prior year's data.

The loader process refers to the movement of the PUR data from the county file into a central DPR database and the identification of errors in the data. One of the most complex parts of the PUR process is the error-handling process, which includes not only the error identification and tracking carried out by the loader program but also the error correction process.

The Importance of Error-handling

Because of the importance of the PUR for many groups and individuals, it is critical that the database be as accurate and complete as possible. People who use the data need to feel confident that the data are sufficiently accurate for their purposes.

Bad data are worse than no data. If a system fails to minimize errors, two serious problems will arise. Some people will use the data not knowing about the errors and, therefore, make wrong conclusions. If laws and regulations are based on these conclusions, there is the potential for serious consequences. Other people will not trust the data and therefore ignore it. In any case, the huge effort at data collection is wasted.

More than 50,000 errors (about 2% of the total number of records) are found each year in the PUR data. Most of these detected errors are corrected before the final version of the PUR is available. However, not all errors can be identified and so the true error rate is unknown. Also, even a 0.1% error rate in pounds applied, if errors are large, could seriously affect some kinds of conclusions.

Although this report describes the error-handling process at DPR, it is important to realize that most error-handling should occur at the time of data entry. If paper reporting forms are used, these forms should be as clear as possible to minimize errors due to misunderstandings. The people filling out the forms should be instructed on the proper procedures. Computer-based forms provide a much more powerful mechanism for preventing errors from being entered because the computer program can check for many possible errors and prevent the entry of invalid values.

Another step in the PUR process is data entry from the PUR reports into the county database. This provides another opportunity for creating errors. Although, the county data entry programs do carry out some error checking routines, many errors are not caught and remain in the data that DPR receives. On the other hand, most county staff

have a good understanding of the PUR process and requirements, and they are invaluable in finding and correcting errors in the reports.

Because of insufficient error-handling procedures in the different county data entry programs, incorrect values can be entered in their databases. Thus, another mechanism must exist at DPR to ferret out these errors. Trying to eliminate and correct errors at this stage in the overall process is much more difficult, much more expensive, and much less successful than preventing errors in the first place. It may be months before county staff find time to investigate the errors, at which time much information may have been forgotten or lost. Also, as time passes people are less motivated to make the corrections.

The difficulties in handling errors at this late stage should become evident from the description of the process in this report.

Overview of the Loader Program

Conventions

This report uses a few conventions to help the reader understand the processes. Database table names are in bold capitals; file names are in italics; the column names in a table are in Arial type. The columns are also called "fields", which should not be confused with agricultural fields, which are always referred to as "agricultural fields". A database record is just a row in a table. As mentioned above, the PUR contains two primary types of records: production agricultural applications and all others. For brevity these are referred to as agricultural applications and non-agricultural applications, respectively.

PUR data loading procedures

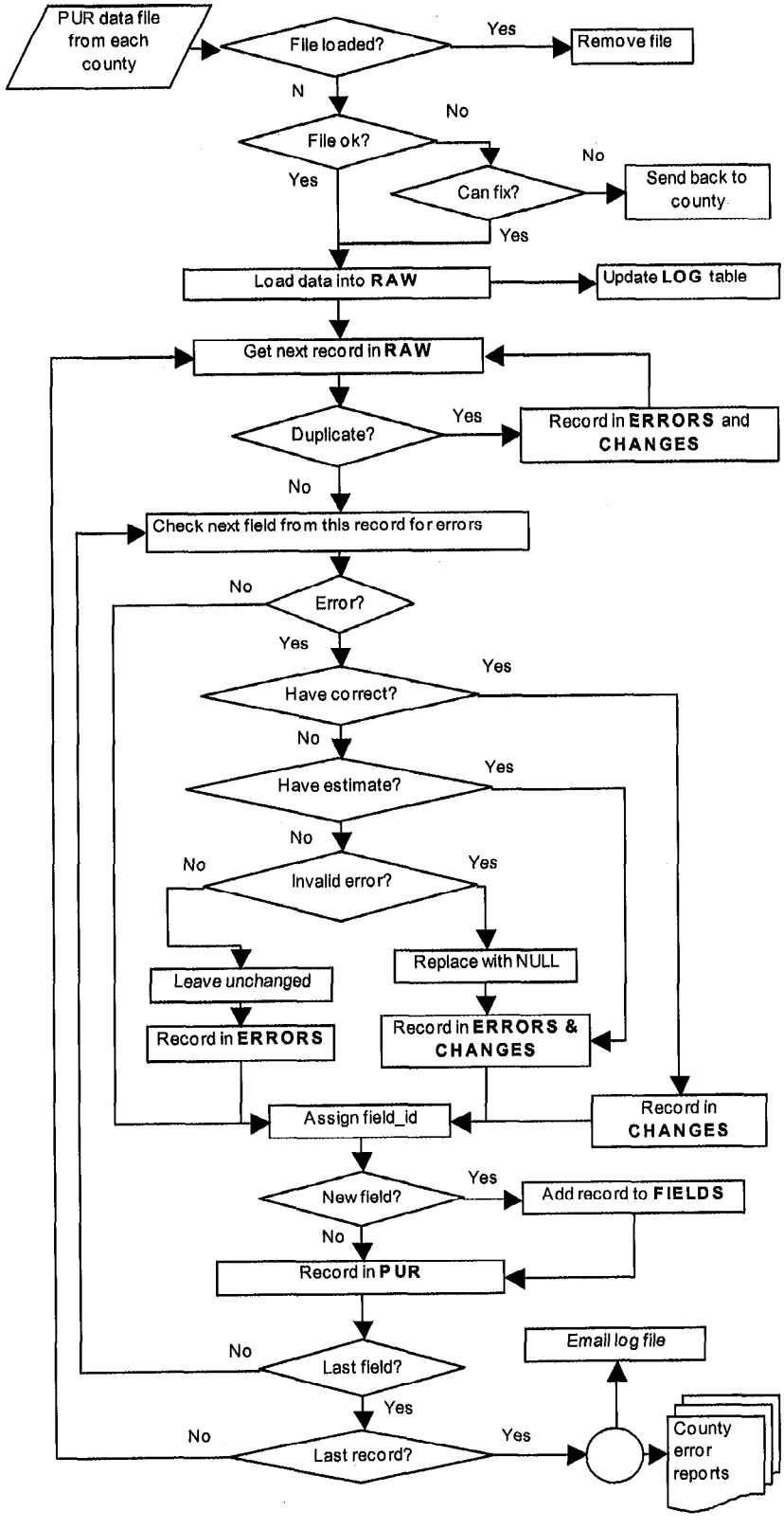
The loader program runs on a UNIX server every day checking for any newly received county PUR data files. When it finds a new file it runs through several steps (Figure 1):

1. checks that data from a PUR data file with the same name as the new file has not already been loaded;
2. checks that the new data file has the correct structure and fixes certain kinds of errors;
3. loads all the data from the new file directly into an Oracle table (**RAW**);
4. logs the name and date of the county file, number of rows of data entered, and the date loaded (**LOG**);
5. checks that each record in the new file is not an erroneous duplicate of records already loaded (not all duplicate records are errors);
6. checks each data field in this table for a series of possible errors; these error checks are described in the next section;
7. where possible, corrects errors or makes estimates to replace invalid data, otherwise replaces the value with a blank or leaves it unchanged;
8. records any uncorrected errors it finds, including both the original value and new value, and the kind of error (**ERRORS**);

9. records any changes made to the data, the date of the change, and whether it was corrected, estimated, or replaced with a null (**CHANGES**);
10. identifies each agricultural field, assigns it a field_id, determines the most likely acres planted and location of the field, and creates a record of this agricultural field (**FIELDS**);
11. if the agricultural field has records in the **PUR** with different values reported for its acres planted or location, the agricultural field is marked as inconsistent in **FIELDS** and a list of all inconsistent values of acres planted and location is made for this field (**FIELD_MTRS_ACRES**);
12. makes some calculations and conversions (such as getting the DPR product number from the pesticide registration number and the pounds of product used from the amount of product used);
13. loads the valid and converted data into another table (**PUR**);
14. creates and prints an error report with all the errors found in the PUR data; and
15. emails loader log files to the loader administrators; these report the data files successfully loaded, the files not loaded because of errors, and any database or operating system error messages that may have been produced during loading.

A more complete description of this process is given in the section "What the loader program does".

Figure 1.Loader program: Overview of procedures to load pesticide use data into DPR database
Database table names in bold capitals



The database tables

The loader program uses, creates, and updates several database tables. It uses several tables from DPR's pesticide label database, such as **PRODUCT**, which lists the names, product numbers (prodno), registration numbers, registration dates, inactive dates if any, and some other information for all California registered pesticide products; **CHEMICAL**, which lists all active ingredients that appear in products; **PROD_CHEM**, which lists the active ingredients in each pesticide product; **SITE**, which lists all the sites that can appear on pesticide labels; and **PROD_SITE**, which lists all sites that appear on the label of each product. It creates and updates other tables, which are described here (and whose fields are described in Tables 1A - D):

PUR: contains all processed and corrected PUR data. It should be used for all official analyses of the PUR data. In order to identify the PUR version being used, analyses should include the date that the table was last updated. This table will also be used to produce annual PUR summary reports. The "preliminary" report will be produced in June and will include all data received and processed by that time. The "final" report will be produced in December and will incorporate the majority of error corrections for that year's data. The December report is called the "final" version because it is the last published report, not because no more errors corrections will be made. This table will continue to be updated as errors are detected and corrected.

RAW: contains original data as it was received from the counties. No changes are made to these data. This table is used to create the other PUR tables and would not normally be used in analyses; previous years' tables would be archived.

ERRORS: contains the current list of all errors found in the PUR data. When values are corrected, the corresponding record in **ERRORS** is deleted.

CHANGES: contains a list of all the changes that have been made to the **PUR** and the time the change was made. Records are never deleted from **CHANGES**, so that it provides a complete history of all changes made to the **PUR**. Thus, this table could be used to recreate an analysis using any particular version of the **PUR**.

LOG: contains a list of all county files that have been loaded into the **PUR**, the number of rows from each file, and the date each file was loaded.

FIELDS: contains a list of each agricultural field in the PUR, identified by the combination of county_cd, grower_id, site_loc_id, and site_code. It contains information about the field such as the acres planted, its location, its region, and a flag that indicates whether different applications on this field reported different values for acres planted or location. This table should be used to calculate agricultural field level statistics and total acres planted.

FIELD_MTRS_ACRES: If a field has inconsistent acres planted or location, records are created in this table listing the field_id, each MTRS and acres planted reported for this

field, and the number of records reported for each combination of MTRS and acres planted. This table would only be used to investigate possible errors in the data.

PUR_SITES: contains all the site or commodity codes and names used in the PUR. This table is a subset of the label database table **SITE,** because not all sites that can appear on pesticide labels are used for use reporting. The field extra contains an 'X' for sites that are not in the list of sites that should be used in the PUR but that were used in some of the earlier years of the PUR. Also, the site names are those used in the PUR explanations given to the counties, which differ somewhat from the names used in **SITE**.

PUR_SITE_QUALIFIER: contains the same sites listed in **PUR_SITES,** but includes the qualifier codes used with each site.

There are a few other tables that are used internally by the loader and error-checking programs. **RAWI, ERRORSI,** and **CHANGESI** are intermediate tables used to temporarily store the data from the current county file that is being loaded and error checked.

Table 1A. All data fields (columns) in the Oracle tables associated with the PUR: **PUR**.

Column name	Type	Description
acre_planted	NUMBER(9, 2)	Size of field that the application occurred on
acre_treated	NUMBER(9, 2)	Size of area or volume treated
aer_gnd_ind	VARCHAR2(1)	Method of application (air, ground, or other)
amt_prd_used	NUMBER(12, 4)	Amount of product applied, units given by unit_of_meas
applic_cnt	NUMBER(6)	Number of pesticide applications for this record
applic_dt	DATE	Date pesticide was applied
applic_time	VARCHAR2(4)	Time of day of application
base_ln_mer	VARCHAR2(1)	Baseline & meridian for application location
batch_no	NUMBER(4)	Number assigned by county to a group of PUR reports
cedts_ind	VARCHAR2(1)	Indicates if report was submitted using the CEDTS system
corr_cntr	NUMBER(1)	Not currently used
county_cd	VARCHAR2(2)	County code for county where pesticide was applied
document_no	VARCHAR2(8)	Number of report in a batch of reports
field_id	NUMBER(8)	System assigned number uniquely identifying each agricultural field
grower_id	VARCHAR2(11)	Value that uniquely identifies each grower or operator
grwr_fut_suf	VARCHAR2(1)	Not currently used
last_up_dt	DATE	Last time any field in this record was changed
lbs_prd_used	NUMBER(14, 4)	Pounds of product applied
license_no	VARCHAR2(13)	License number of pest control operator or business
mtrs	mtrs_type	Meridian-township-range-section value
nursery_ind	VARCHAR2(1)	Not currently used
planting_seq	NUMBER(1)	Number to indicate multiple plantings of a crop on a field
process_mt	NUMBER(2)	Month the PUR report was processed by county
process_yr	NUMBER(4)	Year the PUR report was processed by county
prodno	NUMBER(7)	System assigned pesticide product number
qc_flag1	VARCHAR2(1)	Not currently used
qc_flag2	VARCHAR2(1)	Not currently used
qualify_cd	NUMBER(2)	Code that identifies commodity more precisely than site_code
range	VARCHAR2(2)	Range number for application number
range_dir	VARCHAR2(1)	Range direction for application number
record_id	VARCHAR2(1)	Value indicating the kind of PUR report
section	VARCHAR2(2)	Section number for application number
site_code	NUMBER(6)	Code of site or commodity that was treated
site_loc_id	VARCHAR2(8)	Value that indicates the particular field treated
summary_cd	NUMBER(4)	Line number on PUR report
township	VARCHAR2(2)	Township number for application number
tship_dir	VARCHAR2(1)	Township direction for application number
unit_of_meas	VARCHAR2(2)	Unit of measure for amount product used
unit_planted	VARCHAR2(1)	Unit of area for planted field (acres, square feet, etc)
unit_treated	VARCHAR2(1)	Unit of area or volume treated (acres, cubic feet, etc)
use_no	NUMBER(8)	System assigned number to uniquely identify each record
year	NUMBER(4)	Year of application

Table 1B. All data fields (columns) in the Oracle tables associated with the PUR: **RAW**.

Column name	Type	Description
use_no	NUMBER(8)	System assigned number to uniquely identify each record
record_id	VARCHAR2(1)	Value indicating the kind of PUR report
process_mt	VARCHAR2(2)	Month the PUR report was processed by county
process_yr	VARCHAR2(2)	Year the PUR report was processed by county
batch_no	VARCHAR2(4)	Number assigned by county to a group of PUR data files
nursery_ind	VARCHAR2(1)	Not currently used
county_cd	VARCHAR2(2)	County code for county where pesticide was applied
section	VARCHAR2(2)	Section number for application number
township	VARCHAR2(2)	Township number for application number
tship_dir	VARCHAR2(1)	Township direction for application number
range	VARCHAR2(2)	Range number for application number
range_dir	VARCHAR2(1)	Range direction for application number
base_ln_mer	VARCHAR2(1)	Baseline & meridian for application location
aer_gnd_ind	VARCHAR2(1)	Method of application (air, ground, or other)
grower_id	VARCHAR2(11)	Value that uniquely identifies each grower or operator
cedts_ind	VARCHAR2(1)	Indicates if report was submitted using the CEDTS system
site_loc_id	VARCHAR2(8)	Value that indicates the particular field treated
acre_planted	VARCHAR2(8)	Size of field that the application occurred on
unit_planted	VARCHAR2(1)	Unit of area for planted field (acres, square feet, etc)
applic_dt	VARCHAR2(6)	Date pesticide was applied
site_code	VARCHAR2(6)	Code of site or commodity that was treated
qualify_cd	VARCHAR2(2)	Code that identifies commodity more precisely than site_code
planting_seq	VARCHAR2(1)	Number to indicate multiple plantings of a crop on a field
acre_treated	VARCHAR2(8)	Size of area or volume treated
unit_treated	VARCHAR2(1)	Unit of area or volume treated (acres, cubic feet, etc)
mfg_firmno	VARCHAR2(7)	Manufacturing firm number, component of registration number
label_seq_no	VARCHAR2(5)	Label sequence number, component of registration number
revision_no	VARCHAR2(2)	Revision code, component of registration number
reg_firmno	VARCHAR2(7)	Registrant firm number, component of registration number
amt_prd_used	VARCHAR2(10)	Amount of product applied, units given by unit_of_meas
unit_of_meas	VARCHAR2(2)	Unit of measure for amount product used
document_no	VARCHAR2(8)	Number of report in a batch of reports
summary_cd	VARCHAR2(4)	Line number on PUR report
applic_cnt	VARCHAR2(6)	Number of pesticide applications for this record
applic_time	VARCHAR2(4)	Time of day of application
license_no	VARCHAR2(13)	License number of pest control operator or business
file_date	DATE	Date PUR county file was created
file_name	VARCHAR2(12)	Name of PUR county file

Table 1C. All fields (columns) in the Oracle tables associated with the PUR: **ERRORS, CHANGES, LOG, FIELDS, and FIELD_MTRS_ACRES.**

Fields in ERRORS table

Column name	Type	Description
transaction_no	NUMBER	Uniquely identifies row in errors table
use_no	NUMBER(10)	Foreign key to the PUR tables
error_code	NUMBER(3)	Error code for use by counties
column_name	VARCHAR2(50)	Name of column or field in PUR
old_value	VARCHAR2(50)	Original value in this column (the one with error)
new_value	VARCHAR2(50)	Value used to replace old_value (if do not know correct value or estimate, leave NULL)
duplicate_set	NUMBER(13)	If this record is an error duplicate of other records (error_code = 80), this field gives the number to identify what set of duplicate records this record is in.
error_type	VARCHAR2(20)	"invalid" (value not allowed in this column), "probable" (value allowed but incorrect), "possible" (value may be incorrect), "duplicate" (this row is a duplicate of another one)
replace_type	VARCHAR2(20)	"null" (replace with null) "estimate" (replace with estimate or good guess) "same" (do not replace, leave value same) "delete" (delete entire row),
require_type	VARCHAR2(20)	"required" (value is required for this column) "optional" (value is optional) "not_allowed" (value is not allowed) "unknown" (value may or may not be required)
who	VARCHAR2(30)	Who or which program found the error; if loader, include version number
comments	VARCHAR2(2000)	Comments about the error

Fields in CHANGES table

Column name	Type	Description
transaction_no	NUMBER	Uniquely identifies row in changes table
use_no	NUMBER(10)	Foreign key to the PUR tables
error_code	NUMBER(3)	Error code for use by counties
column_name	VARCHAR2(50)	Name of column or field in PUR
old_value	VARCHAR2(50)	Previous value in this column
new_value	VARCHAR2(50)	New or corrected value
action_taken	VARCHAR2(20)	"null" (replace with null) "estimate" (guess correct value), "correct" (change to correct value), "delete" (delete row), "insert" (insert row), "validate" (indicates that a possible error is correct)
action_time	DATE	Date and time that the tables were updated
who	VARCHAR2(30)	Who made the change; if loader, include version number
county_validated	VARCHAR2(1)	Whether or not the correction was validated by county: Y or N
comments	VARCHAR2(2000)	Comments about the change

Fields in LOG table

Column name	Type	Description
file_name	VARCHAR2(12)	Name of file from county
file_date	DATE	Date of county file
load_date	DATE	Date file loaded
num_of_records	NUMBER(7)	Number of records loaded
start_use_no	NUMBER(8)	First use_no of the records in RAW table
end_use_no	NUMBER(8)	Last use_no

Fields in FIELDS table

Column name	Type	Description
field_id	NUMBER(8)	Unique number assigned by the loader program to identify each agricultural field
county_cd	VARCHAR2(2)	County code
operator_id	VARCHAR2(7)	Operator id, taken from the grower_id
site_loc_id	VARCHAR2(8)	Grower assigned name of the field
site_code	NUMBER(6)	Crop or commodity code
mtrs	mtrs_type	The geographical location of the field using MTRS; if different applications report different values for this field, the value chosen is the one with the largest number of records. Mtrs_type is a user defined type containing the components base_ln_mer, township, tship_dir, range, range_dir, and section. Each component can be identified, for example, by mtrs.base_ln_mer, mtrs.township, etc. You can also get the MTRS value as a single string using mtrs.Get_value().
acres_planted	NUMBER(92)	The acres planted for this field. The unit used is always acres so there is not need for another field for units. If different applications report different values for this field, the value chosen is the one with the largest number of records.
region	VARCHAR2(30)	The name of the region in California where the field is located.
inconsistent_field	VARCHAR2(1)	A flag (an 'X') to indicate that different applications on this field report different acres planted or MTRS.

Fields in FIELD_MTRS_ACRES table

field_id	NUMBER(8)	Unique number assigned by the program to identify each agricultural field. This table lists only inconsistent fields, that is, fields with more than one reported value for acres planted or MTRS. These fields have an 'X' in the inconsistent_field column in FIELDS.
mtrs	mtrs_type	Each MTRS value reported in the PUR for this field.
acres_planted	NUMBER(10,2)	Each acres planted reported in the PUR for this field.
num_recs	NUMBER(12)	The number of records in the PUR with this combination of MTRS and acres planted.

Table 1D. All fields (columns) in the Oracle tables associated with the PUR: **PUR_SITE_QUALIFIER, PUR_SITES**. These tables are subsets of table **SITE** containing only sites that are actually used in the PUR. Also, names are the ones used in the PUR explanations given to the counties.

Fields in PUR_SITE_QUALIFIER table

Column name	Type	Description
site_code	NUMBER(6)	Code for site or commodity that was treated
qualify_cd	NUMBER(2)	Code for additional site specificity
site_name	VARCHAR2(50)	Name of site or commodity
extra	VARCHAR2(1)	An 'X' indicates a site that is not in the list of sites that should be used in PUR reporting but which have appeared in earlier PUR years.

Fields in PUR_SITES table

Column name	Type	Description
site_code	NUMBER(6)	Code for site or commodity that was treated
site_name	VARCHAR2(50)	Name of site or commodity
extra	VARCHAR2(1)	An 'X' indicates a site that is not in the list of sites that should be used in PUR reporting but which have appeared in earlier PUR years.

Error-handling Procedures

Types of Errors and How They Are Handled

The most complex job of the loader program is checking for errors in the PUR data. These errors are classified in a number of ways. **Invalid** errors are values that are not allowed for a field, that is, are not in the set of valid values; for example, a character in a number field, or a unit treated that is not one of the valid values such as A (acres), S (square feet), etc. **Probable** errors are values that are in the set of valid values for a field, but are almost surely wrong, for example, a product registration number that is not in our label database (this could be a correct registration number, the error being in the label database), or an unreasonably high value for the pounds applied. All probable errors, like invalid errors, are reported to the county. **Possible** errors are values that are unusual in some way but not clearly known to be an error, for example, a high value for pounds applied but not so high that it could not have actually been used. Possible errors are not reported to the county, but are left in the **PUR** and **ERRORS** table so they could be identified by anyone wanting to check the data more carefully.

Errors are handled in several ways. (1) If the correct value is known, replace the erroneous value with the correct value. This will probably be an unusual situation, but one example is an error in the first two characters of the grower_id. These characters should be the county code for the county where the pesticide was applied. Since we know the correct county code, we can use that code when there is an error in these characters of the grower_id. (2) If the correct value is unknown but can be guessed or estimated, replace it with the estimate. For example, if the home county code in the grower_id is invalid (does not correspond to one of the county codes) you could assume that it is the same as the county of application, since this is usually true. Another example is estimating an extreme outlier in the rate of use with the median rate for all applications that used this product on this crop or site. (3) If no reasonable estimate can be made and the error is not invalid, leave the value unchanged; this could be considered a kind of estimate. (4) If no reasonable estimate can be made and the value is invalid, replace it with a null. Different characters are used to indicate a null value: '?' is used if the field is a character, -1 if the field is a numeric code for which arithmetic operations are meaningless, and a blank is used if the field is numeric. A '?' cannot be used in a numeric field since it is not a number and '-1' should not be used since it will certainly be an incorrect value and would affect the results of arithmetic operations. An actual character for field and code fields rather than a blank was used to make it easier to link these values to a label such as "UNKNOWN".

An important consideration for determining when values are errors is whether a value is required or not. If a value is not required it may be optional, that is, sometimes reported and sometimes not. In all other cases it is labeled as "not allowed" in the database. Values for some fields are required for all situations, but other values, such as acres planted and geographical location, are required only in certain situations, the most

common one being production agricultural reports (identified by record_id equal to 1, 4, A, or B).

Whenever an error is found and is not corrected, whether the error was invalid, probable, or possible, a record of this is made in the **ERRORS** table. Included in the **ERRORS** record is the use_no, which uniquely identifies the record in the PUR; the data field name; the original value reported that was in error; the value actually entered in the **PUR** table; the error type (invalid, probable, or possible); how the value entered was chosen (by estimation, correction, leaving unchanged, or replacing with null); and whether the value was required, optional, or not required. If the value entered into the **PUR** was different from the originally reported value, even if an error record was not created, a record is made in the **CHANGES** table.

All errors found during error checking are compiled into a printed error report and sent to the appropriate county. County staff make corrections on the printed report and send the marked report back to DPR.

Another program is used by DPR staff to make these corrections to the database. This error correction program calls the same error checking procedures as the loader program to check the new values that staff enter. If a value is invalid, the error correction program will not allow the person to enter the value. If a value is a probable or possible error it will let the person know about the problem, but will allow the person to then override that error check. In some cases values marked as probable errors by the loader program will be found to be correct. The error correction program can be used to change any estimated, incorrect values back to their original, correct values and the record for that error in the **ERRORS** table will be deleted. As with all changes, a record will be made in the **CHANGES** table. The error correction program has not yet been completed and is not described here.

Error Checking Procedures

A description of the errors checked for each of the PUR data fields is given in Tables 2 and 3. Table 2 lists the procedures ordered by error_code and provides a short description of each error; Table 3 provides a more detailed explanation of the procedures for each field in the database. Most of the errors will not be described in this text since Table 3 provides sufficient explanation. However, a few error-checking procedures are rather complex and require further explanation (see below).

Table 2. Error codes and descriptions used for the error reports sent to the counties starting in 2000.

Error Code	Error Description
1	record id is not 1, 2, 4, A, B, or C
2	batch no. is not a number or is 0
3	process month is not a number or not between 1 and 12
4	process year is not a number, is less than 1998, or is greater than 2000
5	county code is not a number or not between 1 and 58
6	section number is not a number or not between 1 and 36
7	township is not a number or not between 1 and 48
8	township direction is not N or S
9	range is not a number or not between 1 and 47
10	range direction is not W or E
11	base meridian is not H, M, or S
12	application method is not A, G, or O
13	commodity code is not a number
14	commodity code is not in commodity table
15	application month is not a number or not between 1 and 12
16	application day is not a number or not between 1 and 31
17	application year is not the current year
18	acres treated is not a number or is 0 where a value is required
19	unit treated is not A, S, C, K, P, T, or U, or is not valid for this type of record
22	area treated is greater than 700 acres or equals 999999 square feet for non production ag
23	area treated is greater than the area of its section
24	registration number (firm) is not a number or is 0
25	registration number (label) is not a number or is 0
30	number of applications is not a number
31	amount used is not a number or is 0
32	unit of measure is not LB, OZ, GA, QT, PT, KG, GR, LI, or ML
34	document number is not a number
35	line item is not a number or is 0
37	product is not in the label database
38	unit of measure is invalid for this formulation
39	commodity is not on the list of commodities for which this pesticide is registered
43	grower id is invalid
44	acres planted is not a number or is 0
45	unit planted is not A, S, C, K, or U
47	area treated is greater than the area planted
48	township-range-section does not exist in this county
51	application date is invalid (e.g. day is not in this month or it is after the current date)
52	specific gravity < 0.0001 if formulation is wet (internal use only)
60	area planted is greater than the area for that section
61	unit treated is not compatible with the unit planted
62	site_loc_id has trailing or leading spaces or missing for production ag record
63	grower_id -site_loc_id-site_code has inconsistent values for MTRS or acre_planted
64	cedits indicator is not E, 0, or space
65	qualify code is not a number

Error Code	Error Description
66	planting sequence is not a number
67	base line meridian-township-range-section does not exist in California
68	application time is not a number or is an invalid time
69	license_no is not valid when no grower id is given
70	pounds of product divided by the acre treated is greater than the first outlier limit
71	pounds of product divided by the acre treated is greater than the second outlier limit
72	pounds of product divided by the acre treated is greater than the third outlier limit
75	rate of use (pounds of product divided by the acres treated) is high
80	record is an erroneous duplicate of another record

Table 3. Description of error checking procedures used in the PUR loader program. The error code is an arbitrary number assigned to each kind of error check. Field is the name of the field in the PUR that is checked. Requirements indicate whether a value is required for this PUR field. Error type indicates whether errors found by this check are invalid, probable, or only possible errors. Replacement indicates whether an error value was replaced with a null character ('?', -1, or null), an estimate of the value, a correct value, or left unchanged.

Error Code	Field	Requirements	Error type	Replace	Error Description
1	record_id	required	invalid	?	Record id is not 1, 2, 4, A, B, or C.
5	county_cd	required	invalid	?	County code is not a number or not between 1 and 58.
13	site_code	required	invalid	-1	Commodity code is not a number.
14	site_code	required	invalid	-1	Commodity code is not in the commodity table.
65	qualify_cd	optional	invalid	-1	Qualifier code is not a number
66	planting_seq	optional for production ag.	invalid	-1	Planting sequence is not a number. Value '0' is replaced with null.
24	mfg_firmno	required	invalid	null	Registration firm number is not a number or is 0
25	label_seq_no	required	invalid	null	Registration label sequence number is not a number or is 0
NA	revision_no	optional	NA	null	Revision code is not between AA and ZZ but no errors are reported. People often do not know the revision code because it is not on some labels. In this case, a default value of AA is used. This may or may not be correct but it is looked at in the check for prodno.
NA	reg_firmno	optional	NA	null	Sub registration number is not a number or is 0. Handled same as for revision code.
37	prodno	required	invalid or possible	-1	Pesticide product is not in DPR's label database. If the program can't find a product with the exact given registration number, it will look for a product with the same 2-part or 3-part registration number and report possible error. If it does not find such a product, it will give an invalid error. If there are several such products, it will choose a product based on several criteria discussed in the loader documentation.

Error Code	Field	Requirements	Error type	Replace	Error Description
39	prodno, site_code	required	invalid or possible	no change	Site code is not on the list of commodities for which this product is registered. This will generate an error message only if prodno is not null, site_code >= 1000, product is not a spreader/sticker, fumigant, rodenticide, or exempt from mill assessment. Possible error occurs where site code is not listed for reported product, but is found for another product with the same 2 or 3-part registration number as reported product.
52	spec_gravity	required	invalid	NA	Specific gravity < 0.0001 if product formulation is wet. Specific gravity value comes from the label database, so it is an error in the label database and is not reported to the counties.
31	amt_prd_used	required	invalid	null	Amount used is not a number or is 0
32	unit_of_meas	required	invalid	?	Unit of measure is not LB, OZ, GA, QT, PT, KG, GR, LI, or ML
38	unit_of_meas	required	probable	no change	Unit of measure is invalid for this formulation; changing this requires that lbs_prd_used be recalculated
18	acre_treated	required if prod ag; required if non-ag with site_code > 100; others optional	invalid	null	Acres treated is null or is not a number or <= 0 when it is required; or acres treated is null and unit treated contains a valid value, when it is optional.
22	acre_treated	required if prod ag; required if non-ag with site_code > 100; others optional	probable	estimate	For non production ag, acres treated is > 700 (unit = A) or acre_treated > 999990 (unit = S) and site is not rangeland, pastureland, forest trees, uncultivated ag, or uncultivated non-ag areas. If error, set unit_treated = U.
23	acre_treated	required if prod ag; required if non-ag with site_code > 100; others optional	probable	estimate	Acres treated is > 10% more than the area of its section, when unit is acres and site is not rangeland, pastureland, forest trees, uncultivated ag, or uncultivated non-ag areas. If error, set acres treated equal to area of the section.
19	unit_treated	required if prod ag; required if non-ag with site_code > 100; others optional	invalid	?	Unit treated is null or is not A, S, C, K, or U (for prod ag) or A, S, C, K, P, T, or U (for others), when it is required. Unit treated is null and acre treated has valid value when it is optional.

Error Code	Field	Requirements	Error type	Replace	Error Description
44	acre_planted	required for production ag.	invalid	null	Acres planted is null or is not a number or <= 0.
60	acre_planted	required for production ag.	probable	estimate	Acres planted is > 10% more than the area of its section, when unit is acres and site is not rangeland, pastureland, forest trees, uncultivated ag, or uncultivated non-ag areas. If error, set acres treated equal to area of the section.
45	unit_planted	required for production ag.	invalid	?	Unit planted is not A, S, C, K, or U.
47	acre_planted, acre_treated	required for production ag.	probable	estimate	Area (or volume) treated is greater than the area (or volume) planted. If error, set area treated equal to area planted. To do the comparison, covert square feet to acres and thousand cubic feet to cubic feet. Special case is if unit_treated = A (or K) and unit_planted = S (or C) and acre_treated < acre_planted, then assume error is with unit_treated and set unit_treated equal to S (or C).
61	unit_treated, unit_planted	required for production ag.	probable, possible, or invalid	estimate, no change, or ?	Unit treated is not compatible with the unit planted, that is, one unit is a measure of area and other is a measure of volume; or one unit is U and other unit is C, K, P, or T.
6	section	required for production ag.	invalid	?	Section number is not a number or not between 1 and 36
7	township	required for production ag.	invalid	?	Township is not a number or not between 1 and 48
8	tship_dir	required for production ag.	invalid	?	Township direction is not N or S
9	range	required for production ag.	invalid	?	Range is not a number or not between 1 and 47
10	range_dir	required for production ag.	invalid	?	Range direction is not W or E
11	base_ln_mer	required for production ag.	invalid	?	Base meridian is not H, M, or S
67	MTRS	required for production ag.	invalid	?	MTRS does not exist in CA
48	MTRS	required for production ag.	invalid	?	MTRS does not exist in this county
43	grower_id	required for production ag; required for non ag. when license_no is null; others optional	invalid	?, estimate, or correct	Grower_id is null, or first two digists not equal to county_cd, or 2nd and 3rd digits not between last two digits of PUR year - 3 and last two digits of PUR year + 3, or 5th and 6th digits not between 1 and 58. Further explanation of the grower_id is given in the documentation.

Error Code	Field	Requirements	Error type	Replace	Error Description
69	license_no	required for non ag. when grower_id is null; optional for non ag. when grower_id has value	invalid	?	License number is null or equals 0.
62	site_loc_id	required for production ag.	invalid if null; probable if extra spaces	? or correct	Site location id is null. If it has trailing or leading spaces, remove the spaces and mark as a probable but corrected error.
63	site_loc_id	required for production ag.	possible	no change	Agricultural field (as identified by grower_id, site_loc_id, and site_code) has inconsistencies. Different applications on a field (identified by the combination of county_cd, grower_id, site_loc_id, and site_code or by the program assigned field_id) report different values for acres planted or MTRS. Further explanation of inconsistent agricultural fields is given in the documentation.
15	applic_dt	required	invalid	null	Application month is not a number or not between 1 and 12.
16	applic_dt	required	invalid	null	Application day is not a number or not between 1 and 31.
17	applic_dt	required	probable	estimate	Application year is not the PUR year. As with process_yr it is converted to 4 digit value. If it is an error, it is estimated as the PUR year.
51	applic_dt	required	invalid	null	Application date is an invalid date (eg day does not exist for the month) or is later than the current date. This error is checked only if there were no errors found for error codes 15, 16, or 17.
68	applic_time	optional for production ag.	invalid	?	Application time's first two digits is not between 0 and 24 or last two digits is not between 0 and 60. Replace any spaces with 0; replace '0000' with null.
30	applic_cnt	optional for non ag.	invalid	null	Number of applications is not a number or < 0 for non ag; replace reported values of 0, blank, or "." with null. For production ag., applic_cnt is always set equal to 1 regardless of value reported and no error report is generated.
12	aer_gnd_ind	required for production ag.	invalid	?	Application method is not A, G, or O

Error Code	Field	Requirements	Error type	Replace	Error Description
2	batch_no	required	invalid	-1	Batch number is not a number or is 0.
34	document_no	required except for CalTrans records	invalid	-1	Document number is not a number. Document_no is not required for reports from CalTrans, so it is set to null in these cases. CalTrans records are identified as records with record_id = 2 and application day = 28.
35	summary_cd	required	invalid	-1	Line item is not a number or is 0.
3	process_mt	required	invalid	null	Process month is not a number or not between 1 and 12.
4	process_yr	required	invalid	null	Process year is not a number, is less than the PUR year, or is greater than 2 + the PUR year. The value from the county file is a two digit number. It is coverted here to a 4 digit year by assuming a number between 50 and 99 is a year in the 1900's and a number between 0 and 49 is a year in the 2000's.
64	cedts_ind	required	invalid	?	Cedits indicator is not null, 0, or E. Value '0' is replaced with null.
75	lbs_prd_used, acre_treated, unit_treated, amt_prd_used	required if prod ag; required if non-ag with site_code > 100; others optional	probable	estimate	Pounds of product divided by the acre treated (the rate of use) is greater than either the first or second outlier limit. Further explanation of extreme rate values is given in the loader documentation.
72	lbs_prd_used, acre_treated, unit_treated, amt_prd_used	required if prod ag; required if non-ag with site_code > 100; others optional	possible	no change	Pounds of product divided by the acre treated (the rate of use) is greater than the third outlier limit. Further explanation of extreme rate values is given in the loader documentation.
80	county_cd, grower_id, site_loc_id, site_code, qualify_cd, prodno, amt_prd_used, unit_of_meas, acre_treated, unit_treated, acre_planted, unit_planted, applic_dt	required for production ag.	probable	delete	Values for these fields are the same as those for 1or more other records and the sum of acres treated for this set is greater than the acres planted. Further explanation of duplicate records is given in the documentation.

Registration number, prodno, and site_code. The person filling out a pesticide use report enters the registration number for the pesticide used. The loader program checks if the given registration number exists in DPR's label database. If not, it will generate an error record.

However, there may be some difficulties in determining which product was used because the submitted form may not give the full pesticide registration number. The United States Environmental Protection Agency (U.S. EPA) pesticide registration number consists of a firm number that identifies the company that is the primary registrant (mfg_firmno), a number assigned to each company's individual product as it is registered (label_seq_no), and a distributor or sub-registrant's number that identifies any company that is marketing a product owned by another company (reg_firmno). In addition, California requires companies to register and license individual brand names, so the California registration number includes an additional code, which uniquely identifies each brand (revision_no). In DPR's database each product with a distinct California registration number is also given a unique identifying number (prodno).

The pesticide label usually displays only the parts assigned by U.S. EPA. If the reg_firmno is different from the mfg_firmno, the label will display the 3-part registration number (mfg_firmno, label_seq_no, and reg_firmno); if the reg_firmno is the same as the mfg_firmno, the label will display the 2-part registration number (mfg_firmno and label_seq_no). The revision_no usually does not appear on the label so it is often not given in the pesticide use reports. If the revision_no is not given, most county data entry programs will provide a default value of 'AA', which may be incorrect. Note that in the county PUR file a reg_firmno value of '0000000' means that reg_firmno is the same as mfg_firmno; in the label database for these cases, reg_firmno will equal mfg_firmno.

To help determine if a reported registration number is valid, the loader program uses information on the reported site treated and date of application. DPR's database includes tables that list all sites that each product is registered for and the dates during which all products had valid registrations. In some situations determining the most likely exact pesticide product is difficult. For example, a reported registration number with a 'AA' revision_no may exist in the database but the reported site may not be on its label but may be on the label for another product that matches the 2 or 3-part registration number. Since a 'AA' revision_no may have been a default value entered by the computer rather than an actual value entered, it could be incorrect. (Of course, it could be incorrect even if someone entered it, but, hopefully, it is more likely to be correct in that situation.)

In this situation, there are several choices:

1. Leave the registration number as reported and report no errors: this choice leaves an unreported error in the PUR (site not on the label) so is not acceptable.

2. Leave the registration number as reported but report the error "site not on label": this does not indicate that there exists a similar product which could be the actual product used so we lose information.
3. Change the value to the registration number with the site on its label and report no errors: this does not give feedback to the county (or users) that there was a problem.
4. Change to the registration number with the site on its label but report site not on label: this is not quite accurate since the site is on the chosen product and it does not distinguish this case from the case where the site code is not on any similar products.
5. Leave the registration number as reported but report a possible error that the site is not on its label and give the registration number of the product with the site on its label.

Choice 5 seems to be the most accurate option and to provide the most information, so was used in the current loader program. There are several other difficult situations, but rather than explain all possibilities, we summarize the algorithm used by the loader:

First check the existence of the reported product registration number:

IF a product is found in DPR's label database with the 4-part registration number,
THEN return the prodno and report no errors
ELSE IF only one product is found in DPR's label database with the three-part registration number reported by the applicator (mfg_firmno, label_seq_no, and reg_firmno)
THEN return the prodno as an estimate and report possible error 37
ELSE IF more than one product is found in DPR's label database with the three-part registration number reported by the applicator (mfg_firmno, label_seq_no, and reg_firmno)
THEN return prodno with a NULL value and report possible error 37 [one of these will be chosen below]
ELSE return prodno with value –1 and report invalid error 37.
END IF.

Second, check site on label and choose a prodno:

IF a prodno was found [it will have a positive value] and the reported site code is on its pesticide label
THEN report no errors
ELSE IF a prodno was found and the site was not on its label
THEN check if site is on the label for other products
    IF site is on any product with the same mfg_firmno and label_seq_no
    THEN report possible error 39 and provide registration number for that product in the comments field
    ELSE report invalid error 39
    END IF.

ELSE IF more than one product was found with the reported 3-part registration number [prodno is NULL]
THEN choose one of them
IF one of the products has a valid specific gravity, has the site on its label, and was registered at the time of application,
THEN report that product as an estimate with possible error 39
ELSE IF a product has the site on its label,
THEN report that product as an estimate with possible error 39
ELSE IF a product has a valid specific gravity,
THEN report that product as an estimate and check other products for site:
IF site is on any product with the same mfg_firmno and label_seq_no
THEN report possible error 39 and provide registration number for that product
ELSE report invalid error 39
END IF
ELSE choose any product in the list, report that product as an estimate, and check other products for site:
IF site is on any product with the same mfg_firmno and label_seq_no
THEN report possible error 39 and provide registration number for that product
ELSE report invalid error 39
END IF
END IF
ELSE [no product was found with the reported 3-part registration number; prodno = -1]
THEN find all the products in DPR's label database with the same mfg_firmno and label_seq_no as that of the reported product and choose one:
IF no product found, THEN report invalid error 39
ELSE IF a product has a valid specific gravity, has the site on its label, and was registered at the time of application,
THEN report that product as estimate with possible error 39
ELSE IF a product has the site on its label,
THEN report that product as estimate with possible error 39
ELSE IF a product has a valid specific gravity,
THEN report that product as estimate with invalid error 39
ELSE choose any product in the list, report that product as estimate with invalid error 39
END IF
END IF

There are a few details in determining if a site_code is on the list of legal sites on the pesticide product label. All sites with site_code < 1000, which are the nursery and mostly non-agricultural sites, are acceptable for any product even if the site is not listed on that product's label. Also, all sites are acceptable for any product that is a spreader/sticker, a fumigant, a rodenticide, or is exempt from mill assessments.

The product is considered to have a valid registration date at the time of application if the time of application is between the original registration date and the date it became

inactive, or after the registration date if the inactive date is null, or before the inactive date if the registration date is null.

Grower_id and license_no. The grower_id is an identifier assigned to the operator or owner of the property where the application occurred. It is composed of: (reporting county_cd) + (year) + (permitting county_cd) + (permit number). The reporting county_cd is the code for the county where the pesticide application occurred; the application year is either the last two digits of the application year, or the expiration year of the pesticide permit, or the year the permit was issued; the permitting county_cd is the code for the county that issued the permit, and the permit number is an arbitrary number assigned to the permit (consisting of five characters). Most counties issue permits each year and in this case the year part of the grower_id should be the year of application. However, some counties issue multi-year permits which can be valid for up to three years. For multi-year permits, some counties set the year part of the grower_id to the year the permit was issued and some counties set it to the year the permit will expire. All counties except for Imperial issue permits in January; Imperial issues them in July. Because of these possibilities, the year part of the grower_id could be any number between the application – 3 to application year + 3.

The license_no is the license number of the person or business who applied the pesticide; the license_no is a new database field added to the PUR in 1999. For production agricultural reports the grower_id is required but the license_no is not required. For non-agricultural reports after 1999, applicators must provide either the grower_id or license_no. They can provide both but that is optional. In 1999, the license_no was considered optional for all reports.

Errors are checked for each of the different parts of the grower_id. If the number in the first two digits of the grower_id is not equal to the county_cd, it is changed by setting it equal to county_cd for the current application. This is not reported as an error, but is recorded in the **CHANGES** table. If the next two digits of the grower_id are not equal to the PUR year – 3 and PUR year + 3, it is also changed to the current year but reported as invalid and as an estimated value. If the value in 5th and 6th place is not a valid county_cd, then it is estimated by setting it equal to the current county_cd. The permitting county could obviously be another county, but in most cases it is the same county so this should be correct in most cases. In any case, this is reported as an error, so the counties have the opportunity to correct it. If the last 7 characters of the grower_id are spaces or all 0, then the grower_id is set to null and is treated as null for all error checking.

No error checking is done for license_no; any set of characters is allowed. However, sometimes an applicator will use the same number for both grower_id and license_no. In this case one of these is likely to be an error so the program assumes that if the number is the format grower_id, it will leave the value in the grower_id field and replace the license_no will null; if the value is not in the grower_id format the program will leave the value in license_no and replace grower_id with null. No error reports are generated, but the changes are recorded in **CHANGES**.

Inconsistent values for an agricultural field. An agricultural field should be identified by the combination of county_cd, grower_id, site_loc_id, and site_code. Since the grower_id contains the county_cd, it is not necessary to include county_cd in this list, but I include it to make the dependency explicit and to provide an important bit of information about the field. However, not all counties or all growers use the site_loc_id in a way to clearly identify different fields, so this can create problems with this interpretation.

If these values do identify an agricultural field, the acres planted and its geographic location should be the same for all reported applications to a given field. Ideally, the PUR database should identify each agricultural field with a unique identifier and contain a separate table with a list of all fields and their unique identifiers, sizes, geographic locations, and maybe other properties. However, the PUR table currently repeats the field properties for each application record. This can result in inconsistencies if different applications to a field are reported with different values for acres planted or geographic location.

At first, during the processing of each county PUR file, the loader would check for such inconsistencies in acres planted and location for each agricultural field. However, this procedure was very slow so the process was changed so that this procedure was done only after all files for one year had been first loaded. The procedure creates and populates a **FIELDS** table. In **FIELDS** the procedure assigns each agricultural field a unique identifier, which is stored in field_id, and stores a set of information about each field.

Location of agricultural fields is given using the Public Lands Survey System of geographical coordinates. These coordinates consist of values for base line meridian, township, township direction, range, range direction, and section. In the PUR each of these components is assigned to a different column. To make handling these values easier, the loader program defines a new user type called mtrs_type which is a composite value containing all these components. This value is referred to here as the "MTRS".

The error checking procedure first determines if the agricultural field reported in the **PUR** record being checked has been identified previously by looking for the field in the **FIELDS** table. If no record for this field appears in **FIELDS**, a new field_id is assigned to the field and a record is added to **FIELDS** containing values for field_id, county_cd, operator_id, site_loc_id, site_code, MTRS, acres_planted, and region. The value for acres planted is assigned only if the reported unit planted was 'A' (acres) or 'S' (square feet). If the units were square feet, the value is converted to acres, otherwise the value in acre_planted is used as is and recorded in **FIELDS**. Thus, acres_planted in **FIELDS** is always in units of acres so there is no need for a units field. **FIELDS** contains one other field called "inconsistent_field" which uses the value "X" as a flag to indicate if any inconsistencies were found for this agricultural field. Since we are discussing here the first record for an agricultural field there can be no inconsistencies and inconsistent_field is left null.

The operator_id is that part of the grower_id which uniquely identifies the operator or grower. The last 5 characters of the grower_id give the operator's permit number. Since different operators in different counties could be assigned the same permit number, to uniquely identify each operator in the state, we need to include the permitting county_cd which are in the 5th and 6th characters of the grower_id. Thus, the last 7 characters are used as the operator_id in **FIELDS**. However, Alameda County has an additional procedure. They will sometimes give one operator two different grower_ids, one that ends in 'R' and another that ends in 'U'. For Alameda County, the operator_id is set equal to the last 7 characters of grower_id, but if there is an 'R' or 'U' on the end, that character is removed.

If the agricultural field for the record being checked had previously reported applications and is, therefore, recorded in **FIELDS**, the procedures checks to see if this field had previously found inconsistencies. If there were no inconsistencies before (that is, the field inconsistent_field is blank) and the acres planted and MTRS for the current record is the same as the previous reported values for this field, nothing else is done. If the current values for acres planted or MTRS are different from the previously consistent values for this field, an 'X' is recorded in inconsistent_field of **FIELDS** and two new rows are inserted into the table **FIELD_MTRS_ACRES**. This table records all the inconsistent values of acres and location for an agricultural field and the number of records in the PUR for each combination of acres and MTRS. We are now discussing the situation of finding the first inconsistent record for an agricultural field, so one of the new rows in **FIELD_MTRS_ACRES** will give the field_id, MTRS, and acres_planted for the current record, and the value 1 for num_recs. The other row will be the same field_id, and the MTRS and acres_planted previously recorded for this field and the number of records in the **PUR** for this agricultural field. Finally, for this situation, an error record will be created in **ERRORSI** for the current record as well as for all the other records in the **PUR** for this agricultural field. These are all flagged as possible errors because we do not know which value is correct. However, the program tries to determine which of the different acres planted or MTRS values are most likely to be the correct values. It skips any missing acres planted values or any MTRS values with missing components and chooses the value that has the largest number of records.

If there were previously found inconsistencies for this field and if the current record has values for acres planted and MTRS that are in the list of previously found values, the record count is increased for this combination of acres planted and MTRS in **FIELD_MTRS_ACRES**. If this combination was not found previously, a new row is added to **FIELD_MTRS_ACRES**. Since the record count has now changed, the program reevaluates the most likely values for acres planted and MTRS as before by choosing the value with the largest number of records and making any changes in **FIELDS**. Of course, a record is added to **ERRORSI** to record the new possible error.

High rates of use. Some PUR records may have extremely large rates of use (pounds of pesticide per area or volume treated). Sometimes unusually large rates are actually applied even when they are larger than the legal maximum rate, but often these values are errors.

How can we determine which values are probably errors? It would help if our database had the maximum label rates, but it doesn't. Even if it did, this would not tell us how large a value must be before it is flagged as a possible error. Several different criteria have been developed to help identify possible errors in rate of use (described in the report "A Computer Program to Identify Outliers in the Pesticide Use Report Database" by Larry Wilhoit, PM 98-01). Three of these criteria were used in the loader program. These criteria are used to flag a record as a possible or probable error if the rate is larger than:

1. 200 pounds of AI per acre treated,
2. the median pounds of product per acre for all uses of that product on a site or commodity, and
3. a value determined by a neural network.

The neural network is the most unusual technique. Basically, it is a computer programming technique to recognize patterns in a set of complex and noisy data. In this case, the distribution of rates have such unusual distributions, standard statistical techniques are of limited use.

The outlier program is run for an entire year of PUR data and creates a table of the outlier rate limits from the different criteria for each combination of product, site or commodity, unit treated, and record type used in the PUR. The loader program uses the previous year's outlier table to decide if each rate value in the file to be loaded is unusually high.

The program treats outliers found by the first two criteria differently from the third (neural network) criterion. If a rate is larger than either limit from the first two criteria it is recorded as a probable error in **ERRORS** and replaced by the median rate of all uses of the product on that site and unit treated and a record of the change is made in **CHANGES**. If it is higher than the third criterion it is left unchanged and recorded as a possible error in **ERRORS**.

Replacing a rate value needs further explanation since there is not a field for rate in the PUR database. A high rate can result from errors in any of five different database fields: unit treated, area treated, pounds of product, amount of product, or the amount's unit of measure. The program goes through the following algorithm for all records with rates of use higher than one of the criteria to determine which value is likely incorrect and changes the value so that it equals or is close to the median rate.

The program first checks the unit_treated. If the unit_treated is not acres, the record is for a production agricultural application, and the site is not commodity or other fumigation, mushrooms, or any of the sites with site_codes between 40010 and 66000 (these include sites such as animals, structures, and aquatic areas), then it recalculates the rate assuming that the units are acres. If the recalculated rate is less than all outlier limits, the loader program changes the units to acres. In the PUR for years after 1994, 97% of all

applications to the sites not excluded used acres for the unit treated, so it is reasonable to assume that in most of these situations the units should be in acres.

For other sites, or where units are already in acres, or where changing to acres still does not make the rate less than the outlier limits, the program then checks the value for acres treated. If this value is less than 1 for any site other than greenhouses, nurseries, any flavoring and spice crop, cilantro, mint, or mustard and if lbs_prd_used/(median rate) is less than acre_planted, then the acre_treated is set equal to lbs_prd_used/(median rate). The median rate is the median rate of use of this product on this site for all applications during the previous year. Making this change to acre_treated will make the rate of use equal to the median rate of use. The above list of sites was excluded because queries revealed that applications are often reported with acre_treated < 1.

For all other situations or where these changes still do not make the rate of use less than the outlier limits, the lbs_prd_used is changed to equal median rate times the acres treated and amt_prd_used is changed proportionally the same. Note that the applicators report only amount of product and unit of measure in their use reports. The loader program calculates the pounds of product from these values.

Duplicate records. Another kind of error that is difficult to handle is an erroneous duplication of an entire record. This could happen, for example, if two different people (say a grower and a PCO) both sent in reports for the same application, if the data for an application was entered twice into the county database, or if the county accidentally sent the same data file twice to DPR. However, some apparent duplicates are correct, for example, if a grower applied 2 or more spot treatments of the same AI to the same field at the same time and reported these as different applications.

Records are considered duplicate if 2 or more records have the same values for county_cd, grower_id, site_loc_id, site_code, qualify_cd, prodno, amt_prd_used, unit_of_meas, acre_treated, unit_treated, acre_planted, unit_planted, applic_dt. The column, applic_time was added in 1999 to help separate error from valid duplicates. However, this data field is not used consistently by all counties so currently is not used in the checks for duplicate records.

The column record_id is not included in the set because the same application could have been reported by a grower (who uses a report with record_id equal to 4 or B) and a PCO (who uses a report with record_id equal to 1 or A). Also, we don't include the MTRS (location) because site_loc_id is supposed to uniquely identify each field for each grower. Errors from using the same site_loc_id for two different fields in different sections were identified as inconsistent field errors (described above). This procedure, of course, will still not find situations where two different fields in the same section had the same site_loc_id.

In order to decide which record duplications are probably errors, the program adds all the acres treated for a set of duplicate records and if this sum is greater than the acres planted for the field, all these records are marked as error duplicates. Not all of these records are

inserted into the **PUR** table. Records from a duplicate set are inserted until the sum of acres treated equals the acres planted. Because this procedure depends on values for acres planted, it can only be used for production agricultural records. Duplicates for non-agricultural records (record_ids of 2 and C) are not flagged as errors.

This procedure, however, will not always correctly identify error duplicates. For example, apparent error duplications would appear if a grower made 2 spot treatments to a field, reported these as two different treatments, but reported the acres treated each time as the entire field. Another example would be two treatments to two different fields that had the same site identification (grower_id and site_loc_id).

We also have to decide what to do if any of these fields are NULL. Since all these fields are required to have a value, the presence of a NULL indicates an error. Most likely, the original erroneous values were the same for each record, so NULLs are considered the same value in each record.

Error duplicate records have error code 80 in **ERRORS**, but to see which records are duplicates of one another, each set of duplicates are given a duplicate set number which is recorded in the column duplicate_set in **ERRORS**. Records that are not inserted into **PUR** are recorded in **CHANGES** with a replace_type value of "delete" to indicate that the entire record was not inserted into the PUR.

Dealing with values that are optional or have unknown requirements. Some field values do not need to be reported but if a value is given it is accepted (such as qualify_cd and planting_seq); these fields are marked as optional. If optional values are given (other than '0' or '.') they are error checked as if they were required; if no values are given (or '0' or '.' which are treated the same as nulls) no error record is created.

Some fields should not have a value (such as site_loc_id for non-agricultural records); these fields are marked as not allowed. If a value is given for fields not allowed, then the value is set to the null character and is not recorded as an error.

In some rare cases, value requirements are unknown. This would happen if the value used to determine requirements (such as the record_id) were unknown. In these cases, the loader decides what to do based on the following algorithm, assuming that the record_id is missing. If a value is optional or not allowed for some record_id values and if a value is required for other record_ids, then if the value is null or 0, record it as a possible error, or if the value is not null but an error, record it as an error. Again, if a value is optional or not allowed for some record_id values, but if no value is required for other record_ids, then if the value is null, do not record it as an error, or if the value is not null but an error, record it as an error.

Calculating Pounds of Product

The loader program during error checking will make some corrections and conversions as described above. However, the most important conversion is calculating pounds of

product used (lbs_prd_used) from the amount of product used and unit of measure as reported by the applicator (Table 4A). In most cases the calculation is simply a matter of converting from one set of units to another. For liquid formulations (reported in volume units such as gallons, pints, liters) you also need the product specific gravity, which the loader gets from the label database. The only difficulty is that the unit of measure of ounces ('oz') has two different meanings. It can refer to either a volume measure or a weight measure. The loader determines which meaning to give ounces by using Table 4B to determine whether a product should be considered liquid or dry based on its formulation, again which comes from the label database. A further difficulty is that pressurized gas formulation for most purposes is considered a liquid (and this is what it is called in the label database), but pesticide applicators typically report amount of pressured gas in measures of weight. Thus the loader assumes ounces of pressured gas measures a weight.

Table 4A. Formulas for calculating the pounds of pesticide product from the reported amount used, for each unit of measure. The specific gravity (density of product/density of water) is needed to convert from volume units to weight units. The density of water is 8.33 lb/gal. 'OZ' can refer to either a liquid or dry measure and this will determine which formula to use. We can determine whether a product is liquid or dry by the pesticide formulation using the relationship in Table 4B.

Unit Meas.	Dry/Wet Meas.	Calculation of lbs_prd_used
LB	D, W	amt_prd_used
OZ	D	amt_prd_used / (16 oz/lb)
GR	D	amt_prd_used / (453.59 g/lb)
KG	D	amt_prd_used * 2.2046 lb/kg
OZ	W	amt_prd_used * specific_gravity * 8.33 lb/gal / (128 oz/gal)
GA	W	amt_prd_used * specific_gravity * 8.33 lb/gal
QT	W	amt_prd_used * specific_gravity * 8.33 lb/gal / (4 qt/gal)
PT	W	amt_prd_used * specific_gravity * 8.33 lb/gal / (8 pt/gal)
LI	W	amt_prd_used * specific_gravity * 8.33 lb/gal / (3.785 li/gal)
ML	W	amt_prd_used * specific_gravity * 8.33 lb/gal / (3785. ml/gal)

Table 4B. Formulation table

Dry/Wet	Formula-tion code	Formulation description
D	A0	DUST/POWDER
W	B0	EMULSIFIABLE CONCENTRATE
W	C0	FLOWABLE CONCENTRATE
W	D0	GEL, PASTE, CREAM
D	E0	GRANULAR/FLAKE
D	F0	IMPREGNATED MATERIAL
W	G0	MICROENCAPSULATED
W	H0	OIL
W	I0	PAINT/COATINGS
D	J0	PELLET/TABLET/CAKE/BRIQUET
D	K0	PRESSURIZED DUST
D	L0	PRESSURIZED GAS
W	M0	PRESSURIZED LIQUID/SPRAYS/FOGGERS
D	N0	SOLUBLE POWDER
W	O0	SOLUTION/LIQUID (READY-TO-USE)
D	P0	WETTABLE POWDER
W	Q0	SUSPENSION
D	R0	DRY FLOWABLE
W	S0	LIQUID CONCENTRATE
W	T0	OTHER (LIQUID)
D	U0	OTHER (DRY)

Computer Program Description

Location of the Loader Program and Output Files

The scripts to load and error check the data are on Environmental Monitoring Branch's UNIX server *pelia*; the script files are all owned by the UNIX user purload and all loader program and output files are in *pelia:/home/purload*. Nearly all directories and files mentioned will be located here. The exception is the directory *scalos:/ora05/data* where DPR staff place the county PUR data files.

Each year's program is in a separate directory in *sql* (such as *sqlpur2000*, *sql/pur2001*, etc.). The program consists of many files, but the main loader program is a perl program, *loader.pl*. This program manages the overall file processing, Oracle and operating system error-handling, and runs several SQL and PL/SQL scripts (with filenames that end in ".sql") and Oracle loader scripts (filenames end in ".ctl"). Most of the other files in this directory are the scripts called by *loader.pl*. The remaining files serve various purposes: *loader.vpj* and *loader.vtg* are created by Visual SlickEdit, which is the programmer editor I used to create the program files; *afiedt.buf* is a temporary file created by SQL*Plus when editing scripts in the buffer; files ending in ".t" are temporary files created from each county file during loading; and ".out" files contain results or error messages from the cron process that runs the loader.

Version 1.0 of the loader program, which was used to load the 1997, 1998, and the preliminary 1999 PUR data are in the directories *sql/pur97*, *sql/pur98*, and *sql/pur99_vers_1*. Version 2.0 of the loader, used for the final 1999 PUR, is in the directory *sql/pur99*. This directory contains two configurations of the program—batch processing and individual county file processing. The batch configuration was used to create the final release of the 1999 PUR and these files are in the directory *sql/pur99/batch*. The county file processing scripts are in the directory *sql/pur99/test*. The loader program will normally be used with the individual county processing and this is the version that is described in this document. The batch configuration was only used after DPR had received most or all of the county files, so these were all loaded at once, rather than file by file. The files in *sql/pur99/test* were used to test the new version, and will be implemented starting in 2000.

All county files and files produced by the loader program are in the directories *pur1997*, *pur1998*, *pur1999*, *pur2000*, etc.. The directory *pur2000* and for later years contains several subdirectories. Directory *pur2000/data* contains copies of the county files; *pur2000/load_logs* contains log files of the loading process; *pur2000/loaded* contains the county files that were successfully loaded; *pur2000/loading* contains county files that are in the process of being loaded; *pur2000/not_loaded* contains files that could not be loaded because of some kind of errors; and *pur2000/reports* contains the county error reports that are sent to the counties. Directories for the other years contain similar kinds of subdirectories.

This document and other documentation files are in *docs/loader*.

Monitoring the Loader Program
The main program, loader.pl, is started automatically by a cron process each day at a certain time. You can see the time it will start by logging onto UNIX as user purload and typing: "crontab –l". The first two numbers of the output gives the min and hour of the day that *loader.pl* will start. You can change this by editing the cron file using "crontab –e".

After each day's run, the loader will email one or two log files to the list of loader administrators. One log file gives a more detailed description, results, and error messages, if any, from each major step in the loader process. The other log file simply gives a list of county files found, a list of the files that were successfully loaded, and a list of files that were not loaded because of errors.

If no problems appeared, then the loader process needs no other attention. If there were problems, the detailed log file will explain the problem and its location (for example, which script and where in the script). This may involve checking either the county file for problems or the loader script.

To set up a new year's loading process, you need to create the necessary directories and copy the code to the new directories. To help in that process, you can use the perl script *setup.pl*

What the Loader Program Does
A complete explanation of what the program does is found in the source code. These files are extensively commented to help a programmer understand the code.

Before the loading process starts, the loader checks several things. First it checks that the program is not still running from a previous time; if it is still running it will not start another process. If there are large PUR county files or many files to process, the program may take longer than one day to run. If it is not already running, it will open a log file to record the program's progress and any errors than may occur. This file is saved into the directory *pur2000/load_logs* and will have the current date and time appended to the filename. The loader then checks that all needed files are present and will quit if any are missing. The log file will contain a list of the missing files. It will then check that the compiled procedures are up to date; if not, it will recompile them from the necessary files. Finally, it will check if any new PUR files have been copied to the designated directory. If so, it will first copy the files (for backup) to *pur2000/data* and then move the files to *pur2000/loading*. At this point, the normal PUR file processing can begin.

The loader program will first get the name of each PUR county file in *pur2000/loading* and examine the county file's basic structure. First, it creates a temporary copy of the file, and in that copy it replaces a string of one or more control characters with the UNIX end of line character. It also finds the number of characters in each line of the file and the number of lines in the file. If the line lengths are different it will print error messages giving the line lengths for the different line numbers and will not load that file. Finally, it

adds the name of the file and last modification date of the original PUR file to the end of each line in the temporary copy.

Next, if the file contains more than 15,000 lines, it will split the file into nearly equal sized smaller files and name each file with the name of the original file with "_1", "_2", etc. added to the end of name. The reason for splitting large files is that the program runs inefficiently for large files. It usually takes more than one day to process files with more than 15,000 lines; for files much bigger, it could take the program weeks to finish. However, the same data distributed in several smaller files would be loaded much sooner.

It next checks if data from a file with the same name was loaded previously. If a file with the same name was previously loaded, the program will not load the data from the new file. Instead it will move the file to *pur2000/not_loaded*, generate an error message, and then go to the next file in *pur2000/loading*. If the file was not previously loaded, the loader will then check if the intermediate tables (**RAWI, ERRORSI,** and **CHANGESI**) are ready to process another file; unless something went seriously wrong with the loader program these tables should have no data because they only store data temporarily while processing each file.

Next, the loader calls SQL*Loader to load all the data in the file into the Oracle table **RAWI**. This table consists of character fields, so the data gets loaded exactly as it is in the file. There are four common types of files: one contains production agricultural records, another non-agricultural records, and, for each record type, there are the old file formats (before 1999 with no fields for license_no and applic_time) and the new file structures (starting in 1999 with the two new fields). The loader still needs to check for old file formats because not all counties have switched to the new format. There is, also, a special format for files from CalTrans, which have their own unique structure. The loader determines which file format to use based on the number of characters in each line. After SQL*Loader runs, it creates a log file with information and possible errors messages. These files are saved in the directory *pur2000/load_logs/oracle*. Also, if the data were successfully loaded, a record is inserted in **LOG** with the name of the file, number of lines loaded, and the current date. The loader checks that the number of new rows in **RAWI** equals the number of lines in the PUR file found previously when the loader checked the file structure.

The next check is the error checking of the PUR data, now loaded into **RAWI**. The code for this is in two compiled PL/SQL packages. One package, Co_error, contains one primary procedure that calls separate procedures from another package, Check_value, that checks for errors in each field of the PUR. The Check_value procedures take as input parameters the use_no or the current record and the value of the current field being checked. Some checks require other information such as the record_id or county_cd. Each procedure returns values for an error_code, error_type, replace_type, and require_type. The error_code will return a number corresponding to the error found, or, if no errors are found, it will return a null. The error_type can be 'invalid', 'probable', 'possible', 'duplicate', or 'N' (for no errors). The replace_type can be 'null', 'estimate', 'same', or 'delete'. The require_type can be 'required', 'optional', 'not_allowed', or 'unknown'.

The purpose and meaning of each of these terms is explained above in the section "Types of Errors and How They Are Handled". The procedures in Check_value only checks for errors.

The procedures in Co_error use the Check_value procedures to decide how the data will be handled. Values with no errors will be recorded into the table **PUR**. If an error is found, a null character, an estimated value, a corrected value, or the original value will be recorded into **PUR** depending on the parameters returned about from the Check_value procedures. Also, if errors were found and not corrected a record will be inserted in **ERRORSI** and if the value entered in **PUR** was in anyway different from the original value in the PUR file, a record will be inserted into **CHANGESI**.

Because each field has its own separate procedure, these procedures can also be used in the later stages of the PUR system when corrections need to be made. Staff use another program to make corrections to the PUR database. This error correction program calls the Check_value procedures to determine if new values entered by staff contain errors.

After each row from the PUR file is checked for errors, the program assigns the next sequential number to the use_no value, which is then stored in **RAWI** and **PUR**. After all data are error checked, the starting and ending use_no values for this file are then recorded in **LOG**.

An error report is then created from the data in **RAWI** and **ERRORSI** and saved in the directory *pur2000/reports*. The loader program prints the error report and staff mail the report to the county that produced the original data file. This report contains a list of each line in **RAWI** (which is the same as a line in the PUR file) that contains at least one error and, for each line, a list of all the errors found in that line. It also includes a table listing all product registration numbers and site_codes where the site_code was not found on the list of commodities for that product; a table listing each error found with the number of records that had that error; the number of invalid and valid records and the total number of errors for each record_id; and the total number of records for each record_id, application month, and batch_no. Since the loader program runs every day and prints an error report for each county file, the error reports normally are sent back to the counties after a few days from the time the county files were received at DPR.

The loader then moves all records from the intermediate tables to the permanent tables, moves the PUR file in *pur2000/loading* to the directory *pur2000/loaded*; and deletes the temporary copy of the county PUR file.

Finally after all the PUR files have been processed, one of two log files are emailed to a list of people interested in monitoring the loading process. One log file contains a list of all PUR county files found, files that were successfully loaded, and files were not loaded because of errors. The other log file contains the same information but in addition it contains a detailed record of the loader progress during each major step of the process and all error messages, if any, that were generated during the process (such as out of memory errors, missing code scripts, or problems in the structure of any data files).

Rationale for the Program Structure

The processes described here may seem excessively complex, but there were clear and conscious reasons why it was designed this way. The database and error tracking tables need to be flexible and complete enough for several different purposes. Relating the history of the loader program may help in understanding why the program was created in the way it was.

The loader program was originally written by the Information Systems Branch using the database program FOCUS. In 1999, the responsibility for the PUR was divided between the Enforcement Branch and the Environmental Monitoring and Pest Management Branch (EMPM), with Enforcement taking responsibility for the interaction with the counties and EMPM processing and storing the data. Also, the database system was changed from Focus to Oracle. In the process of rewriting the loading and error-checking procedures for Oracle, we had the opportunity to improve the system and add other error-checking procedures.

Based on my experience with the PUR data I suggested a few changes and additions to the error checks that were previously done on the data. These suggestions were sent to several DPR staff and were discussed at a meeting of these staff on March 3, 1999. A description of the agreements we made on the error checks are given in the file *doc/new_validations.doc*. Appendix 1 describes the major differences between the error checks done previously with the checks described in this documentation.

Responsibility for rewriting the loader program was originally given to Yihua Lin. When she left the department, the responsibility was moved to Steve Kishaba. However, Steve was later moved off that project, which was then handed over to Larry Wilhoit. The code Yihua and Steve developed is in the directory *sql/yihua_steve_khoas_files*. They had made a good start, but did not finish it.

Yihua's approach to error checking was to write separate SQL scripts for each field. These scripts would scan through the entire PUR table for each field and record any errors found in another table. For simple error checking this worked fairly well, but more complicated error checking required the use of a procedural language. So I rewrote the routines in PL/SQL. This also vastly improved the efficiency because rather than scanning the entire PUR table every time for each column, the PL/SQL code read each row or record into memory and checked all values in that row for errors. Thus, the PUR table was scanned through just once. Also, many of the fields depend on or are affected by values in other fields. In the SQL scripts these other fields would have to be read in separately again (another costly operation) while the PL/SQL code had all values for a row already in memory. Another difficulty with the SQL scripts was that if any unexpected errors occurred the entire script was stopped. The PL/SQL code can handle unexpected errors in appropriate ways. Finally, the PL/SQL structure made the code much easier to understand.

However, my first attempt at creating this code (which are the ones I have used for loading the 1998 and 1999 data and are in the directories *sql/pur98* and *sql/pur99_vers_1*) had a number of problems. First, it was written as one function and as it got longer with additional and more complex error checks, it became difficult to follow especially when conditional statements spanned several pages. It was difficult to know where each statement was in relation to the various conditions and difficult to follow the logic.

Also, this function was written to both load the data into Oracle tables and check for errors. However, error checking should also be done when any changes are made to the PUR. The error checking routines in the loading code could not be used as they were since they were embedded in code for another function. So, to make the code easier to follow and to separate the error checking function from the loading function, yet another rewriting of code was started and this became version 2.

This latest version creates separate PL/SQL functions for each value and each function only checks for errors. These functions require several input values and several output values that can be used for different functions. The input values include the value from the county file as a character and other values from the current record as necessary, such as record_id, to determine what error checks were appropriate. The output values included the correct value in the correct type, or null value if it was invalid, or the correct or estimated value, and the error code and type of error if there was an error, whether the value was corrected or estimated, and whether the value was required.

In addition, version 2 improved on many of the error procedures previously written and added some additional error checks that I never had time to do in the first version, the most important ones being duplicate records and inconsistent agricultural fields. Version 2 was also extensively tested for programming bugs. Finally, the Oracle table structure was improved, including more information on errors found, and made it easier to use and understand. For example, in version 1 both errors found and changes made were recorded in one Oracle table. These two very different kinds of records are now kept in different tables.

The **CHANGES** table is important to track the history of changes to the PUR which occur through the year and even after the "final" PUR is released. Thus, queries run at different times may use slightly different versions of the PUR. It may be important to know which version of the PUR was used, which can be done by reporting the last PUR update along with the data from the query. If for some reason a query from an earlier PUR version needs to be replicated, a person can use the **CHANGES** table to recreate the data from a previous version of the PUR.

Conclusion

All PUR data received from all counties in California are now processed by the loader program described in this document. The loader program runs every day and processes all the files received from the counties, checks the data for errors, loads valid data into the PUR table, flags any errors it finds, and generates an error report that is sent back to the appropriate county for correction. The original data received from the counties is kept in a separate database table. The loader program also maintains another table that lists all the errors found and another table that keeps a record of all changes that have been made to the PUR. All these tables can be viewed or queried by DPR staff.

This program adds several improvements over the previous loader program:

1. It adds several additional error checks thus improving the accuracy of the data;
2. It leaves out of the database tables only invalid values (or replaces them with estimates) not entire records that contain an error in any of their fields;
3. It includes records for illegal applications;
4. It is continually being updated and corrected so that queries will always have the most accurate data available;
5. It contains a table listing all current errors or possible errors so that staff can see what problems may exist in the data;
6. It contains a table giving the history all changes made to the database;
7. It contains a table listing all the county files received and the number of lines from the file that were successfully loaded;
8. It automates many of the loader tasks, possessing data and emailing the results of the process every day without the need for staff intervention;
9. The program and how to manage the process is fully documented;

Appendix 1: Differences Between the New and Previous Error Procedures

This new error checking procedure has several significant differences from the procedure used in previous years. One difference is that all errors found and all changes made are recorded and tracked in database tables. This allows people to see which values have errors and the entire history of which values have been corrected or changed.

In contrast to the previous procedure where, if a value was found to be an error, the entire record for that application was not entered into the PUR, the new procedure enters the record but changes (usually by making it null) only the fields that contain errors. Thus the data will be more complete and accurate. Another significant difference is that previously any error found that was a legal violation (such as applying a pesticide to a crop that was not on the label) the record of that application was not entered. With the new procedure these records are entered into the PUR database, again making the data more complete.

Finally, several new error procedures were added and previous ones improved. The primary additions are the checks for extremely high rates of use, agricultural field consistencies, duplicate records, but there are several others. Further details of the differences can be seen by comparing the Appendix Table, which lists the error procedures carried out in 1997, and Table 2, which lists the new error procedures. The Appendix Table includes a column that flags error procedures that differ significantly from the new procedures. These are the only descriptions or explanations DPR supplied for the error checks and it is not always clear what some of these descriptions mean or to what kinds of records they applied to. In some cases they appear to differ from what was in the actual code.

Appendix 1 Table. Error codes and descriptions used in the loader program in 1997. The column "Differ" contains an "X" if the new error checks differ from these. The new error checks also include several additional error checks.

Error Code	Differ	Error Description
1		record id not numeric or equal 1, 2, 4, 9, A, B, C, D
2		batch no. not numeric or less than 0001
3		process date (month) not numeric or not between 1 and 12
4	X	process date (year) not numeric or not between 90 and 99
5		county no. not numeric or not = 01 thru 58
6		section no. not numeric or not = 01 thru 36
7		township no. not numeric or not = 01 thru 48
8		township direction not equal N, S, or space
9		range not numeric or not = 01 thru 47
10		range direction not equal E, W, or space
11		base meridian not equal H, S, M, or space
12		application method not equal A, G, or O
13		commodity code not numeric
14		commodity code not on commodity table
15		date applied (month) not numeric or not = 01 thru 12
16		date applied (day) not numeric or not = 01 thru 31
17		date applied (year) not equal 97
18	X	acres treated not numeric
19	X	units not equal A, C, K, P, S, T, U or space
21	X	units = A, C, K, P S, T, or U and units treated equal zero
22	X	units = space, and units treated not equal zero
23	X	units = A, acres treated greater than 700.00
24		reg. no. (firm) not numeric, or equal zeros
25		reg. no. (label) not numeric, or equal zeros
26	X	reg. no. (rev code) not numeric, or equal spaces
27		reg. no. (sub reg) not numeric
30	X	number of applications not numeric or equal zeros
31		amount used not numeric, equal zero
32		unit of meas. not = lb, oz, ga, qt, pt, gr, kg, li, ml)
34		document no. not numeric or equal spaces
35		line item not numeric
37		product not on master label file
38		error in lbs conversion (check formulation)
39		commodity not on master label file
40	X	units not A, C, K, P, S, T, U or units treated = zero
41	X	units not space or units treated not zero
43	X	invalid grower id
44		acres planted must be numeric
45	X	invalid unit planted, see 19
47	X	treated units greater than planted units
48		invalid combination of county + township + range + section
50	X	amount product used > 3000 and unit of meas is GA
51		invalid application date (i.e. 04/31/97 or > current date)
52		check spec. gravity & formulation (internal used only)