Using Pesticide Use Report Data

Larry Wilhoit
Department of Pesticide Regulation
California Environmental Protection Agency
Topics of Discussion

- How to get the data
- How to get information from the data
- Using SAS
- Using MySQL, an open source database
- Using Excel’s pivot tables
- Conclusion
How to Get PUR Data

- DPR’s PUR Annual Report
  (www.cdpr.ca.gov/docs/pur/purmain.htm)
- CDROM of entire database: 3.5 million records/year, 30 data fields, in 58 county text files
- California Pesticide Information Portal (Cal/PIP)
  (calpip.cdpr.ca.gov/cfdocs/calpip/prod/main.cfm)
How to Get PUR Data
How to Get PUR Data

[Image of the California Pesticide Information Portal]

[Description of the portal:
- **Department of Pesticide Regulation**
- **California Pesticide Information Portal**
- **Query Pesticide Use Report**
- **Available Data Columns:**
  - YEAR
  - APPLICATION_MONTH
  - MONTHLY_YEAR
  - DATE
  - COUNTY_NAME
  - COUNTY_CODE

- **Output Columns Selected:**
  - YEAR
  - APPLICATION_MONTH
  - DATE
  - COUNTY_NAME
  - COUNTY_CODE

- **Columns to Sort By:**
  - UNIT_TREATED
  - AG_NEIGH
  - SITE_LOCATION_ID
  - GROWER_ID
  - LICENSE
  - USE_NUMBER

- **Sort-By Columns Selected:**
  - SITE_NAME
  - CHEMICAL_NAME

- **Options:**
  - Summarize the Data
  - Format Output
  - Help

[Instructions for using the portal]

[Note: The portal is used to query pesticide use data, and the screenshot shows how to select and sort data columns for a report.]

[Version: 2009.11 (2009 PUR Data Update)]
How to Get PUR Data

Getting data from Cal/PIP is preferable to the CDROM:

- You get one large tab-delimited text file per year rather than 58 comma delimited files
- You can get just the subset of the data you need
- You can get the variables you want, denormalized
- You get the latest, corrected data
- However, not all fields are available
How to Get PUR Data

- UCIPM web site
  (www.ipm.ucdavis.edu/PUSE/puse1.html)
- Pesticide Action Network (PAN)
  (www.pesticideinfo.org/Search_Use.jsp)
- Department of Health Services (www.ehib.org)
How to Get PUR Data

Research and IPM

Research Tools: California Pesticide Use Summaries

The information in this database was provided by the California Environmental Protection Agency, Department of Pesticide Regulation. The database includes summaries of pesticide use by site, pesticide, county, and month for 1980 through 2000 only. No other information about a pesticide or its label is available in this database.

This form allows you to submit database requests to the UC IPM computer for processing and generation of a report or data file. You request dates, counties, sites, and pesticides and specify report or data file formats.

CAUTION: Specify only the criteria you need. Large requests increase file size and retrieval time.

Pesticide use period:

- Pesticide use period:
  - 2000

From: January Through: December
How to Get PUR Data

PAN Pesticides Database - California Pesticide Use

Pesticide Use in California

California pesticide use data show that between 1991 and 2000 almost 2 billion pounds of active ingredients were applied in California alone. After a massive increase in pesticide use in the early to mid-1990's, reported use has stabilized at about 200 million pounds of active ingredients each year. This figure only includes farm use and professional pesticide use. Not included are consumer and much institutional pesticide use. Also not included in this figure are so-called ' inert' ingredients. U.S. pesticide use is about 1.2 billion pounds each year, and worldwide pesticide use is about 5 billion pounds each year. For detailed information on pesticide use in the U.S. overall and in the California, New York or Oregon pesticide use reporting systems, please see our Pesticide Use pages.

Search CA Pesticide Use


Find pesticide use information on over 250 crops and sites in California for 2004. Crops are organized into the following groups:

- Vegetables and Melons
- Fruits and Nuts
- Field Crops
- Spices and Herbs
- Nursery Products
- Forests
- Livestock
- Other Agriculture
- Non-Agricultural

CA Crop Use

CA County Use

Additional Resources

Find pesticide use information for 58 California counties for 2004.

Links to additional information on pesticide use in California, other states, the U.S. and Europe.
How to Get PUR Data

Pesticide Use in California 1991-2004 - Microsoft Internet Explorer

PAN Pesticides Database - California Pesticide Use

Pesticide Use in California

Note: See Working with the Information on this Page section below for important notes about this data.


All Chemicals (Chem Code 00, All used on Figs (Site Code: 600) in California (County Code: 00)
Acres planted estimated by PUR data for figs appear to be fairly accurate. Low acreage in 1992 is probably incorrect due to errors in the original data.
Note: Data for All Chemicals summarizes data from 2 or more chem codes and may not agree with DPR summaries which total data for each chem code.

Onerous Pounds: 43,918 33,831 54,234 36,906 26,654 26,428 49,064 38,656 31,981 34,768 83,591 45,377 40,689 34,341
Apil Count: 640 528 526 723 752 603 655 617 448 488 592 660 497 465
Field Count: 1 90 101 120 114 99 118 136 113 109 109 99 92 86
Acres Plant: 120.0 8,820 10,420 9,833 10,452 10,113 11,148 12,864 11,166 12,189 11,003 10,431 9,525 9,990
Apil Rate: 0.50 0.37 0.90 0.52 0.36 0.40 0.92 0.65 0.76 0.64 2.06 0.66 1.22 0.92
How to Get PUR Data
How to Get PUR Data

- Special request to DPR

- Reasons for special request:
  - You need data on variables not available in web interfaces (num apps, formulation, reg no, CAS number)
  - You need special formatting of some values (such as dates or NULLS)
  - You need more complex or unusual kinds of queries and do not have database skills

- Ask for the data in the appropriate format for your needs
An Example: Subtropical Fruit

Get PUR data for subtropical fruits from 1992 to 2005: avocado, banana, date, fig, kiwi, mango, persimmon, pomegranate

Data on year, crops, application date, application_type, AI, pesticide product, registration number, formulation, crop, pounds AI, acres treated, acres planted, county, MTRS, grower_id, site_loc_id
An Example: Subtropical Fruit

- For the demonstration, I created a data file using Oracle SQL
- The result is a tab delimited text file where
  - The first line is a list of the variable names
  - It contains data from years 1992 to 2005
  - It has almost 280,000 records
How to Get Information from the Data

- What do you do with such a large data file?
- Excel is limited to 65,000 records (Excel 2007 to 1 million records)
  - If file is not too large you can split it into several files
- Importing into a database is a better option
Databases

- Microsoft Access
  - Relatively inexpensive
  - Many people familiar with it
  - However, cannot gracefully handle very large files

- Major databases: powerful but expensive
  - Oracle
  - Microsoft SQL Server
  - IBM’s DB2
Databases

- Statistical packages: e.g. SAS
- Open source databases: powerful and cheap
  - MySQL
  - PostgreSQL
- All these databases use the SQL database language
- But you can use ODBC to view the data in other programs
Using SAS

- **SAS advantages**
  - Can handle very large files
  - Runs fast
  - Has many tools for analyzing data, including both SQL statements and SAS procedures
  - The main advantage is its set of statistical procedures
Using SAS

Importing data into SAS

- SAS import procedure
  - Easier to use
  - May not import some data correctly
  - Use default parameter values, except may need to set QUESINGROWS = 30000

- SAS datastep
  - More difficult to use
  - Allows you to specify data type
Finding the total pounds of AI on all subtropical fruits each year from 1992 to 2005

This query is done both with a SQL script and a SAS procedure

You can also send the results to an Excel file
Using SAS

- Now calculate a series of statistical measures for rates of use by each year, crop, and AI.
- Finally, graph the frequency distribution of the log rate of use for glyphosate on avocados.
Using an Open Source Database

- Two important databases are MySQL and PostgreSQL
- These are free and have all the power most people need
- To get MySQL go to dev.mysql.com/downloads/
- To get PostgreSQL go to www.postgresql.org/download/
Using an Open Source Database

- I will use MySQL in this demo
- See PUR training docs at agis.ucdavis.edu/pur for installing MySQL
- Access MySQL data using a command line interface
Using Open Source Database

- To import tab delimited text file replace NULLs with “\N”, dates in ANSI format “YYYY-MM-DD”
- To import data, first create a database, then create a table, then import data into the table.
- If you know SQL you are ready to go
- Otherwise set up the MySQL ODBC driver
Viewing MySQL data from Excel

- Access the MySQL data using the ODBC connection from the menu “New Database Query…”
- To see data in easier format use pivot tables
Using Excel Pivot Tables

- Pivot tables are interactive crosstabs
- They allow you to easily summarize the data in many different ways
- But watch out!
More about the PUR

- PUR workgroup web site:
  [http://agis.ucdavis.edu/pur](http://agis.ucdavis.edu/pur)

- This site will have my presentation today plus a Word file with scripts and further explanations.
Conclusions

- There are many different software tools that can be used to more easily get the information you want.
- Software demonstrated today included SAS, MySQL, ODBC, and Excel.