Pesticide Use Reporting System (PUR)

Larry Wilhoit

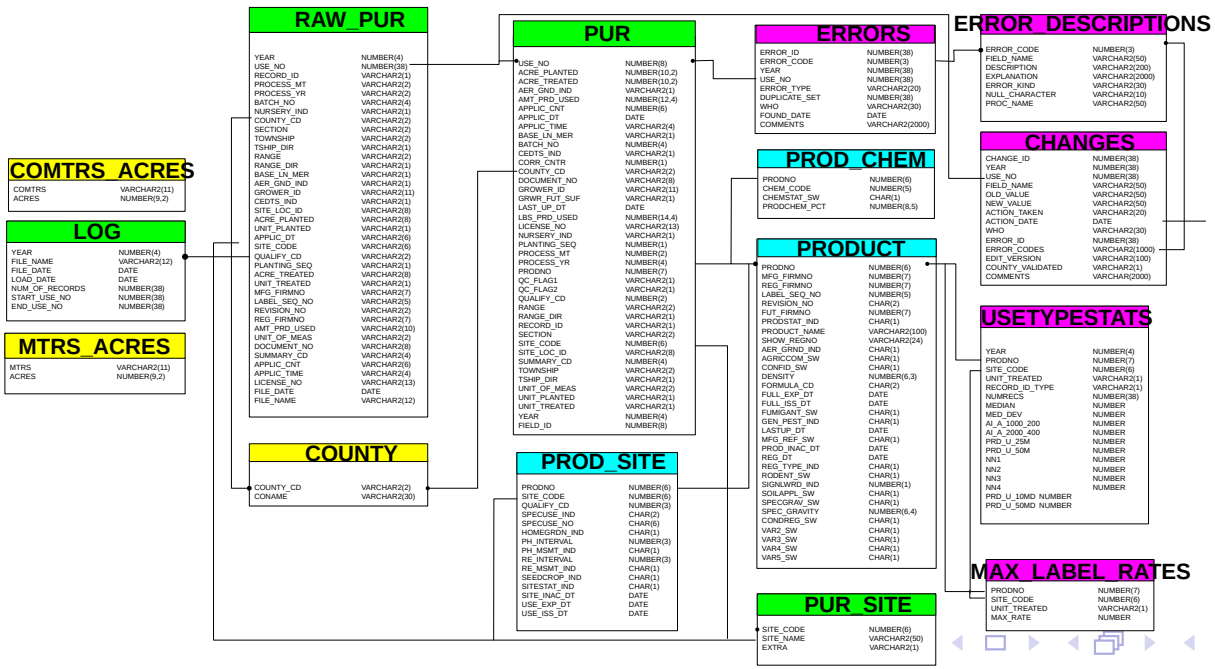
August 4, 2016

California Environmental Protection Agency
dpr Department of Pesticide Regulation

Outline

The PUR has appeared in three different forms

- Limited reporting from 1934 to 1956.
- Reporting of restricted pesticides and applications by commercial applicators from 1970 to 1989.
- Reporting of all production agricultural applications and commerical applications from 1990 to now.



California Environmental Protection Agency
Department of Pesticide Regulation

When did the PUR start?

- The beginnings of the PUR seem to be lost in the mists of time.
- Several DPR documents state: “California has had limited pesticide use reporting since at least 1950.”
- What is the source for this statement?

Sources of statements on earliest reporting

From DPR's *Guide to Pesticide Regulation in California*

"County agricultural commissioners required agricultural pest control operators to send monthly reports. County requirements varied but many included a statement for each application showing the grower's name, location, treatment date, crop, acres or other units treated, target pest, kind of pesticide used, and the strength and amount of the pesticide applied."

From Department of Agriculture's *Monthly Bulletin*, 1950

"County agricultural commissioners require agricultural pest control operators to submit monthly reports of their work. The requirements of counties differ, but many require a statement with regard to each application, showing the grower's name, location, spraying dates, crop, acres or other units treated, pest, kind of pesticide used, and the strength and amount of the spray or dust mixture applied."

The Department of Agricultural Bulletins

- The Department of Agricultural bulletins included tables with pesticide use for each year, starting with 1934.
- These tables continued until 1956 (except 1944 and 1945).
- Pesticide use was reported in acres treated by airplanes.

The Department of Agricultural Bulletins

- Tables in 1954 and 1955 also included applications made by ground. Did these include all agricultural use?
- Reports usually included tables with use by chemical, crop, and pest. Later included use by county.
- Why were these data collected? CAC regulations?

The start of the 1970 - 1989 PUR

The "real" PUR started in 1970 as a result of FAC 14012, created by Statutes of 1969, Chapter 1169

".. each commissioner shall report to the director the type and amount of injurious materials for which permits are issued pursuant to Chapter 3 (commencing with Section 14001), Division 7, the crop and pest or pests on which the material is used, and any other relevant information the director may require. The contents of these reports shall be summarized quarterly by the director as to the type of material and amounts, and the summaries shall be made a public record, providing the name of the registrant filing a report shall be confidential."

The start of the 1970 - 1989 PUR

The regulations of 1973 determine who should report and what should be reported

"Each person licensed for hire in the business of pest control shall maintain a record of the following information . . . and shall report to the commissioner: 1) date and time of the treatment, 2) owner or operator, 3) location of property, 4) commodity or thing, 5) total acreage, 6) pest to be controlled, 7) material and dosage used, 8) equipment used, 9) person who applied the treatment, 10) temperature, 11) wind direction and velocity, 12) other information required by the commissioner. "

Start of the current PUR

The current PUR started in 1990 as a result of Food Safety Act of 1989 (Chapter 12001, AB 2161)

".. A pesticide use report shall be submitted to the commissioner or director on a form and in a manner prescribed by the director. The data from the pesticide use reports shall be considered in setting priorities for food monitoring, pesticide use enforcement, farm worker safety programs, environmental monitoring, pest control research, public health monitoring and research, and similar activities by the department, or by the department in cooperation with other state, regional, or local agencies with appropriate authority."

PUR significant events

- 1990. The 1990 data had numerous problems and are therefore less reliable that later years. The 1990 site codes used were different than those used in later years.
- Early 1990's. PUR records with errors and with illegal uses were not entered into the database. These records were kept in another table, but that table was lost.
- 1994. DPR helped create an electronic system for submission of use reports, called the California Electronic Data Transfer System.
- 1997. The PUR Annual Report for the first time included a section on trends of use.

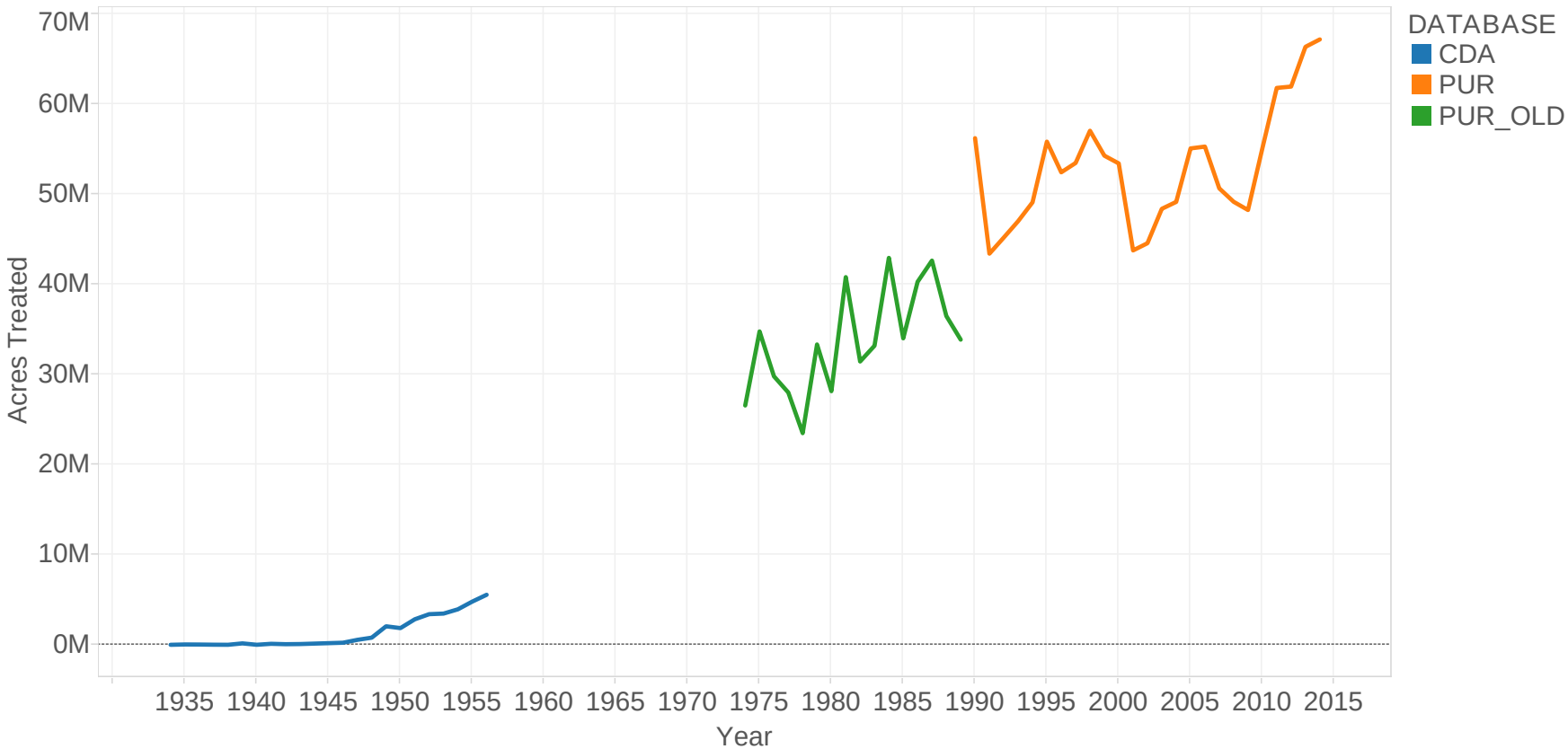
PUR significant events

- 1999. Current PUR loader program was implemented. Earliest year with data for the RAW_PUR, ERRORS, and CHANGES tables.
- 2000. PUR Conference held to educate people about the data and to get suggestions on how the database could be improved. Attended by over 100 people.
- 2002. PUR workgroup formed and started annual meetings with invited speakers to present their work involving the PUR.

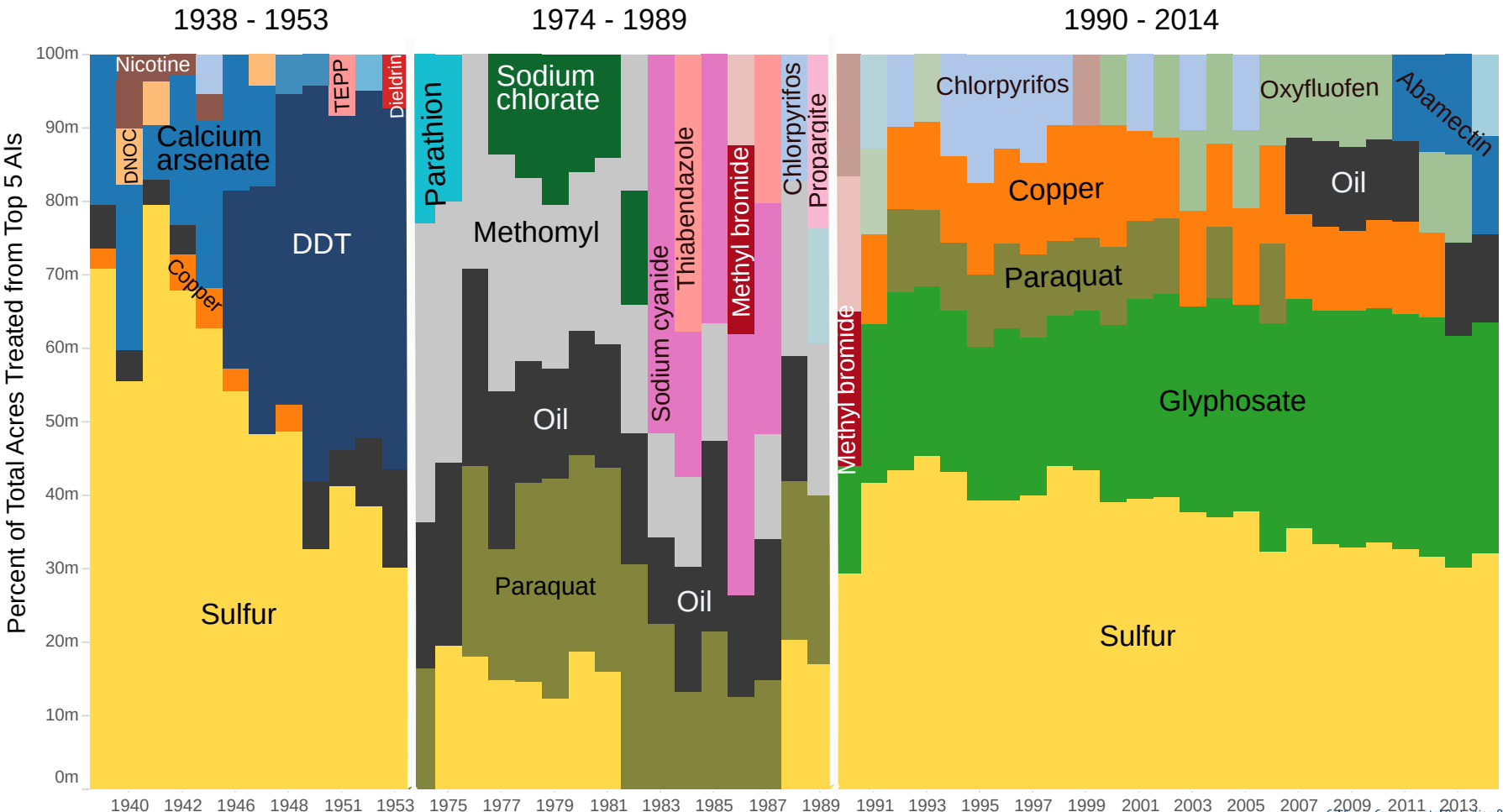
PUR significant events

- 2002. DPR started collecting use information from businesses that apply pesticides at California public K-12 schools and licensed childcare centers. These data are stored in a separate database.
- 2003. DPR launched the Web-based California Pesticide Information Portal (CalPIP), which provides interactive access to the PUR.
- 2011. CalAgPermits created by the County Agricultural Commissioners with support from DPR. This is a web based system for issuing pesticide permits and entering and transmitting PUR data. This greatly improved the error reporting process.

Total acres treated each year from 1934 to 1956 and 1974 to 2014



Top 5 AIs from 1938 to 1953 and from 1974 to 2014



Errors in Amount Used or Area Treated

- Errors in amount or area can be found most easily by unusually high or low rates of use
- Difficulties in finding errors in rates of use
  - Rates vary by AI, product, site treated, target pest
  - May need to search millions of records
  - Rates often have complex frequency distributions

General Methods of Detecting Outliers

- Maximum label rates
- Fixed limits, such as 100 pounds per acre
- Compare the rate to rates of previously reported applications with similar kinds of use

Current Outlier Criteria - Used Since 1998

- A rate of use is considered an outlier if it is greater than:
  - Fixed limit: 200 lbs/acre for most AIs or 1000 lbs/acre for fumigants
  - Median of rates of similar use times 50
  - Median of rates of similar use plus MAD times 50
  - Limit determined by neural networks (not used after 2009)

Similar uses are all applications in the previous year with same product, site, record type (ag or non-ag), and unit treated.

Ways to Improve the Criteria

- For the fixed limits, create more groups of AIs than just fumigants and all others
- Define similar uses as applications of each AI rather than each product and site
- Use rates from last 5 years rather than just last year
- Use log of rates, rather than just rates, in defining criteria
- Develop better robust measures of dispersion
- Develop limits for all non-agricultural records using pounds per application

Grouping Similar Rates

Problem

For many AIs there are two or more distinct groups, each group with very different rates. How can we identify such groups?

Possible methods

Two possible algorithms, Regression Trees and Patient Rule-Induction Method (PRIM), are methods that partitions data into sets based on values of different predictor variables. They differ in how they determine the sets.

Grouping Similar Rates

Possible methods

Regression trees try to make the response average in each set as different as possible. It creates separate groups if the total sum of squares is much larger than the sum of squares of each group.

Possible methods

PRIM tries to create separate groups where the difference between the trimmed means of the groups is sufficiently large.

Result

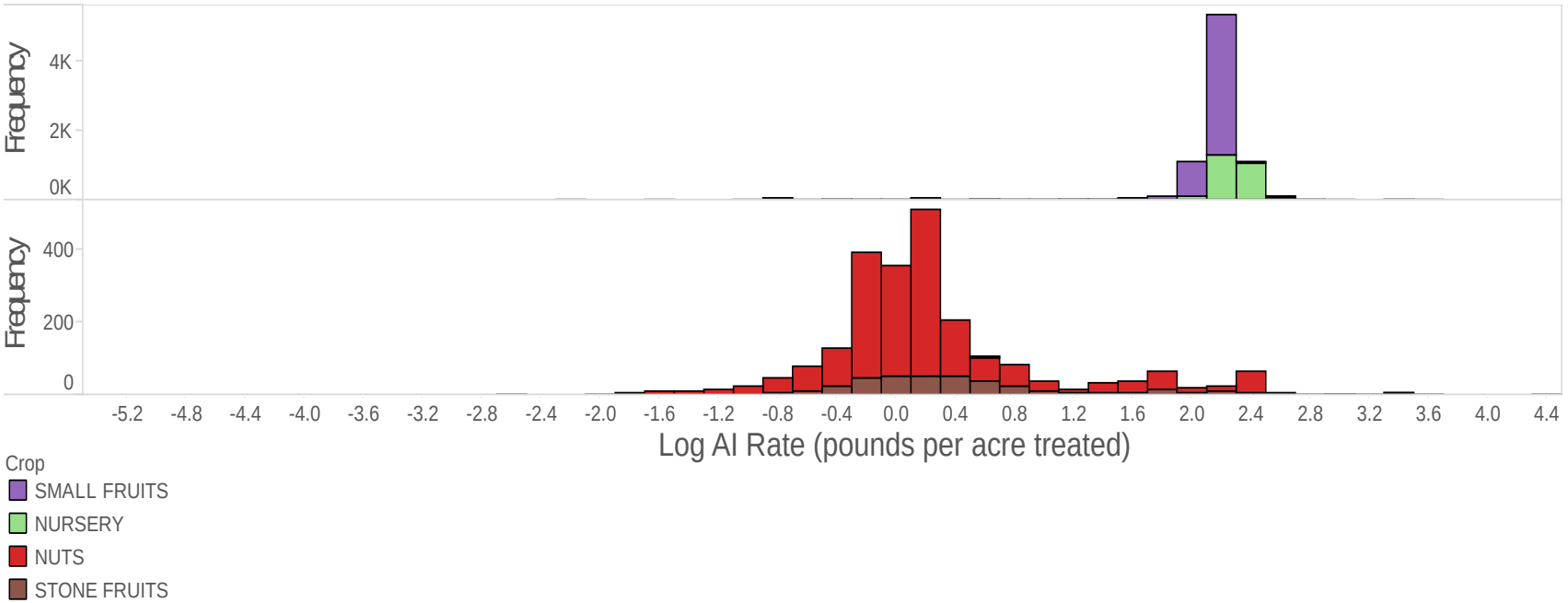
Neither procedure works well for finding distinct AI rate groups.

Grouping Similar Rates

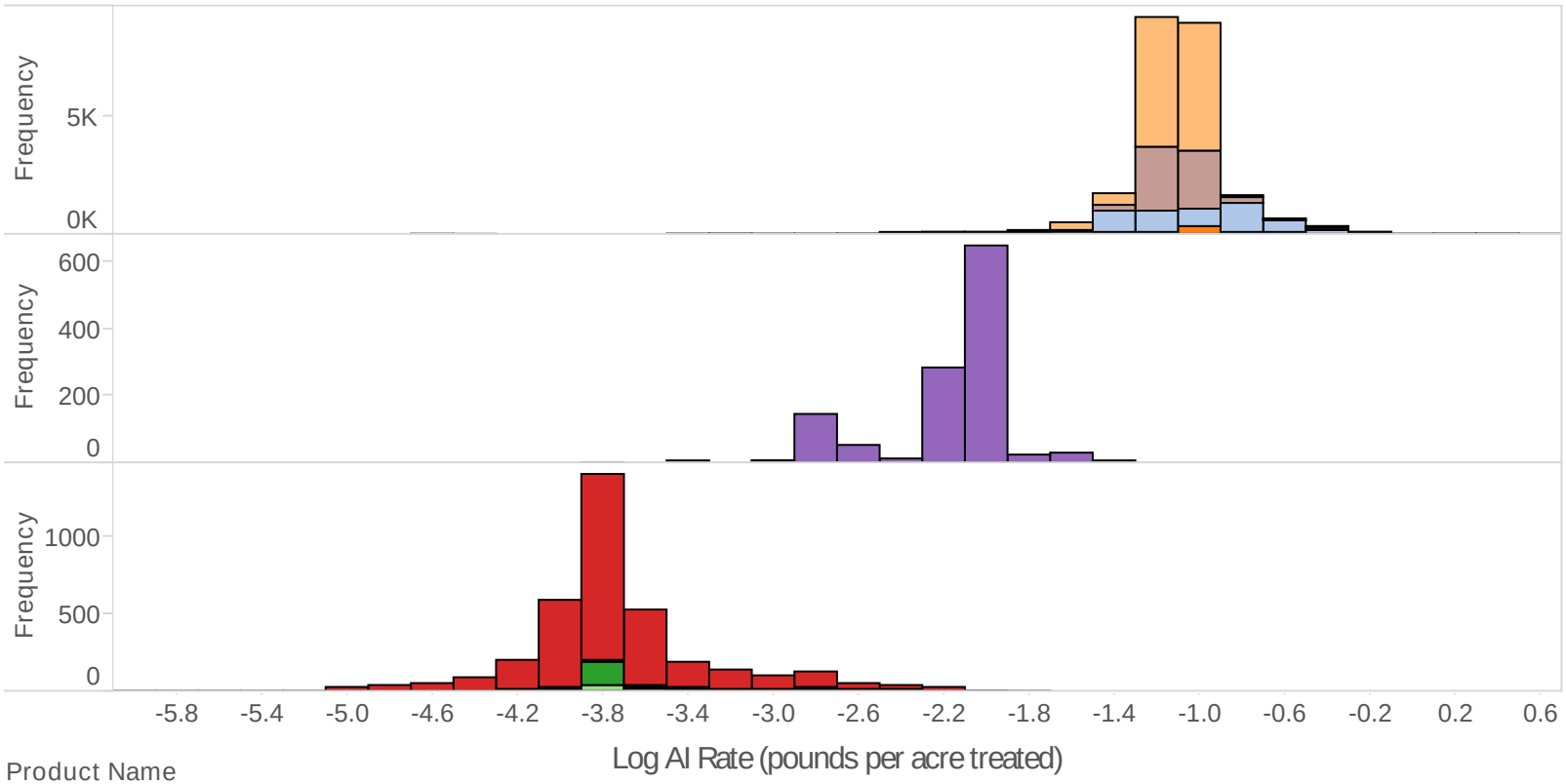
New grouping algorithm

A new algorithm combines ideas from both of these procedures. It creates separate groups if both the difference between the trimmed means is large and the total sum of squares is much larger than the sum of squares of each group. Groups are determined by looking at all combinations of sites treated and products used.

Algorithm Applied to Methyl Bromide Rates



Algorithm Applied to Spinosad Rates



Proposed New Outlier Criteria

- New criteria use different limits for different kinds of applications, based on:
  - Record type - agricultural using pounds per unit treated or non-agricultural using pounds per application
  - Unit treated - acres, cubic feet, pounds, or miscellaneous
  - AI rate type - high rate, medium rate, normal rate, or adjuvant
  - Site type - water site or anything else
- Two kinds of criteria are used:
  - Fixed value for each type of application
  - Trimmed mean + C $\times$ SD: Trimmed mean log (rates of similar use) + C $\times$ trimmed standard deviation, where C is a constant
- For normal rate AIs in ag, the recommended outlier limits are:
  - Fixed limit at 100 pounds per acre and
  - Trimmed mean limit at trimmed mean + 10 $\times$ standard deviation.

Proposed New Outlier Criteria

Similar uses are applications with same AI, AI group as determined by the grouping algorithm, record type (ag or non-ag), and unit treated.

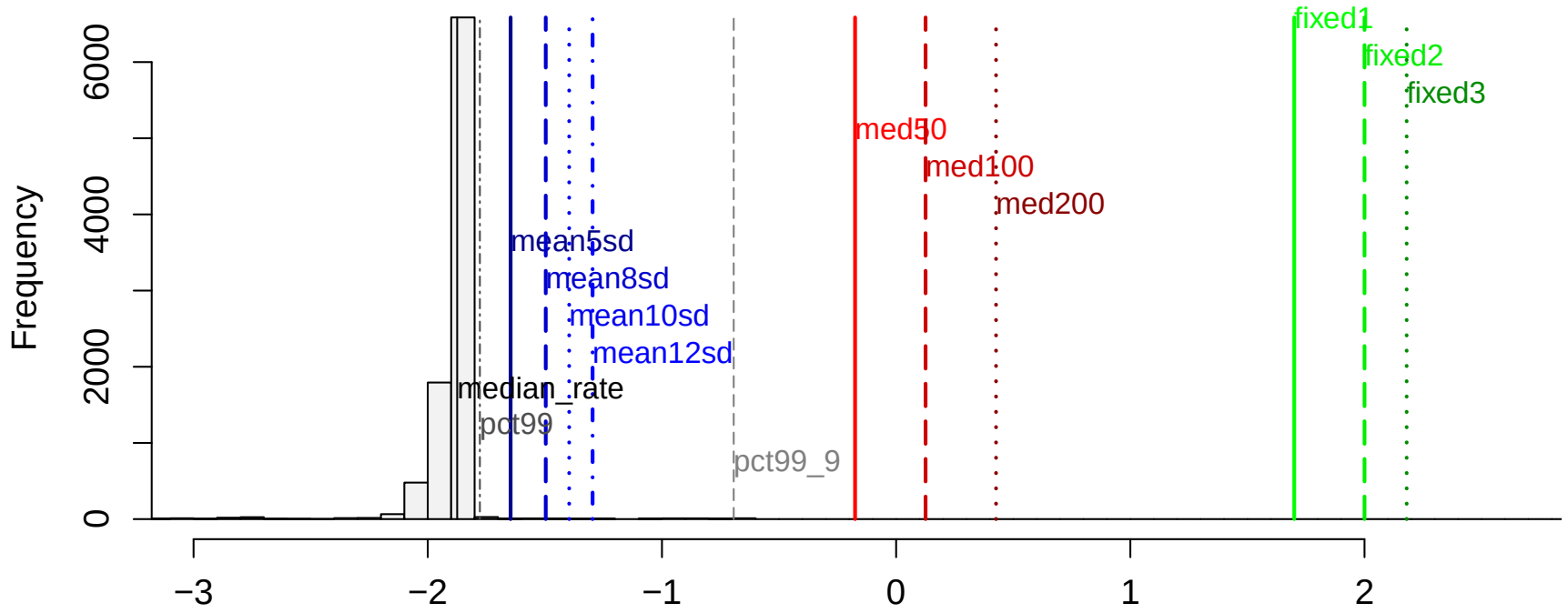
- No longer need median times some constant. If median limit $>$ trimmed mean limit then record will be flagged by trimmed mean; median would be useful if there were any outliers $>$ median limit and $<$ trimmed limit — but such cases are extremely rare.
- No longer use neural networks. Vast improvements in this area (deep learning) make this a potentially very good method.

Criteria Evaluation

1. Visual inspections of distributions with the different outlier limits
2. Percent of records identifed as outliers by each criterion
3. Applying criteria to data with known maximum label rates

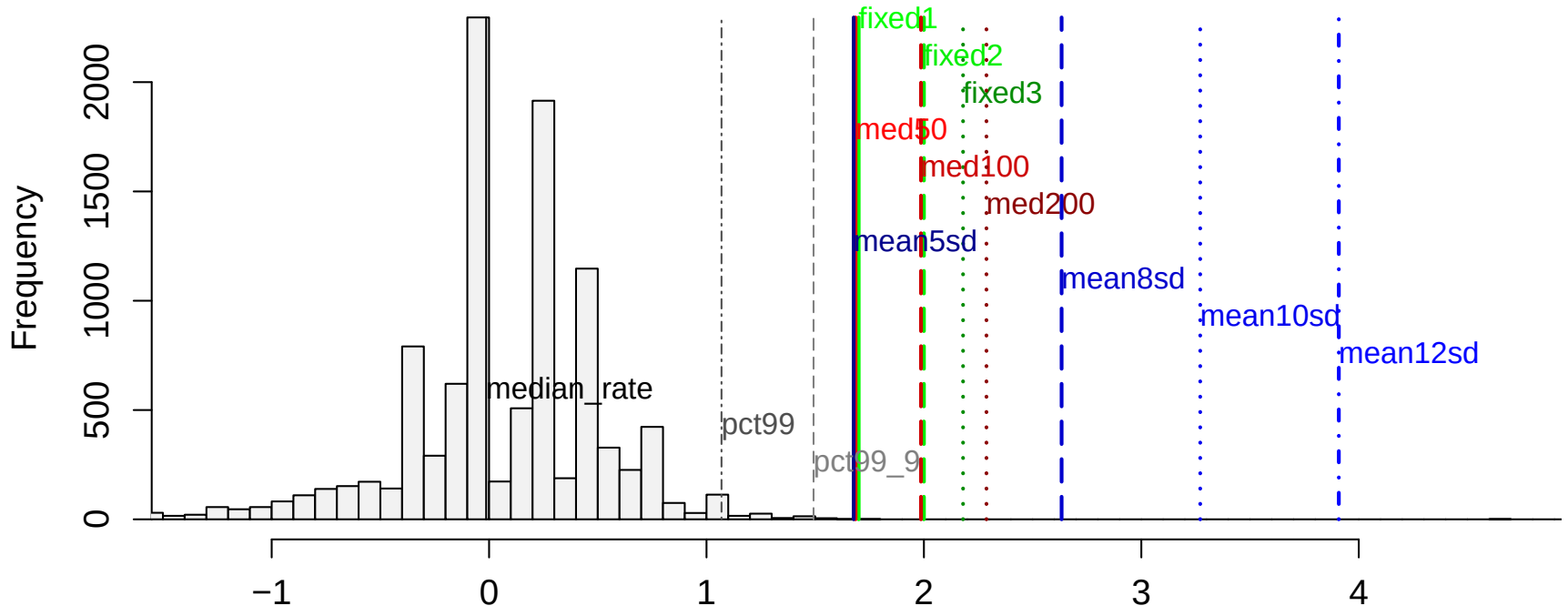
Visual Inspections of Distributions

Mesosulfuron-methyl: pounds of AI per acre



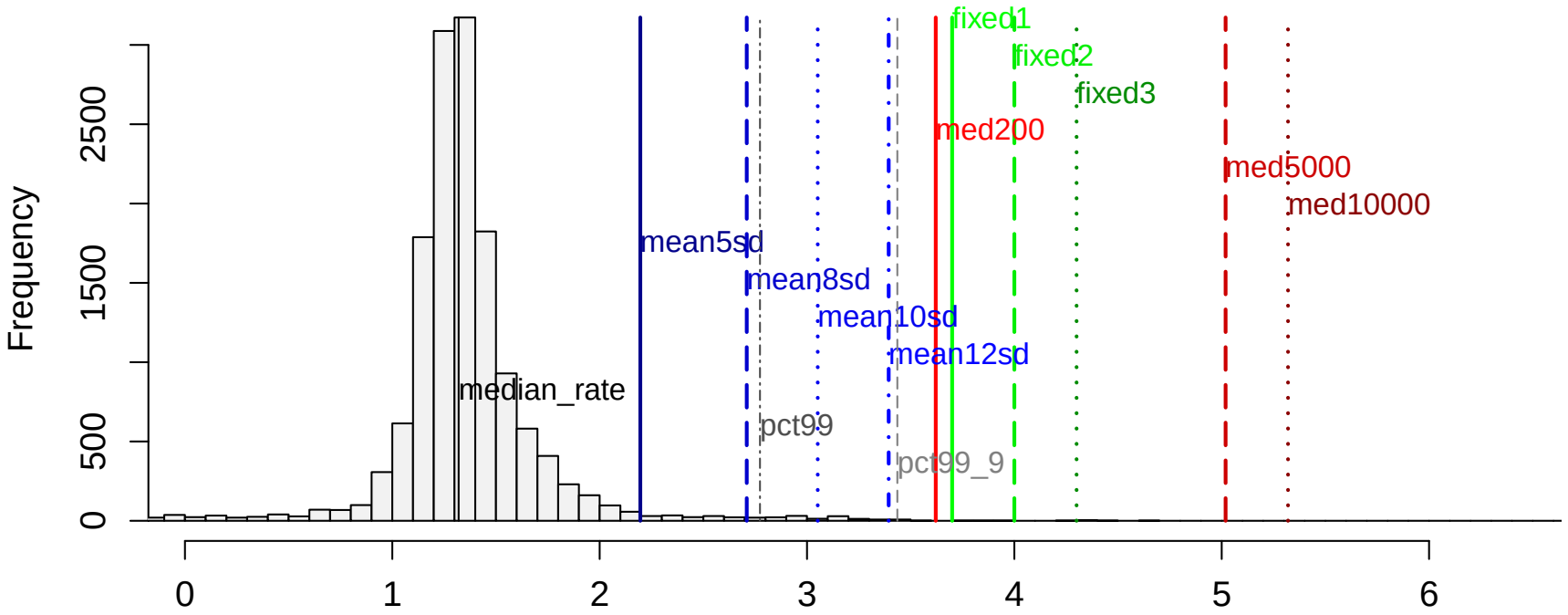
Visual Inspections of Distributions

Diquat dibromide: pounds of AI per acre



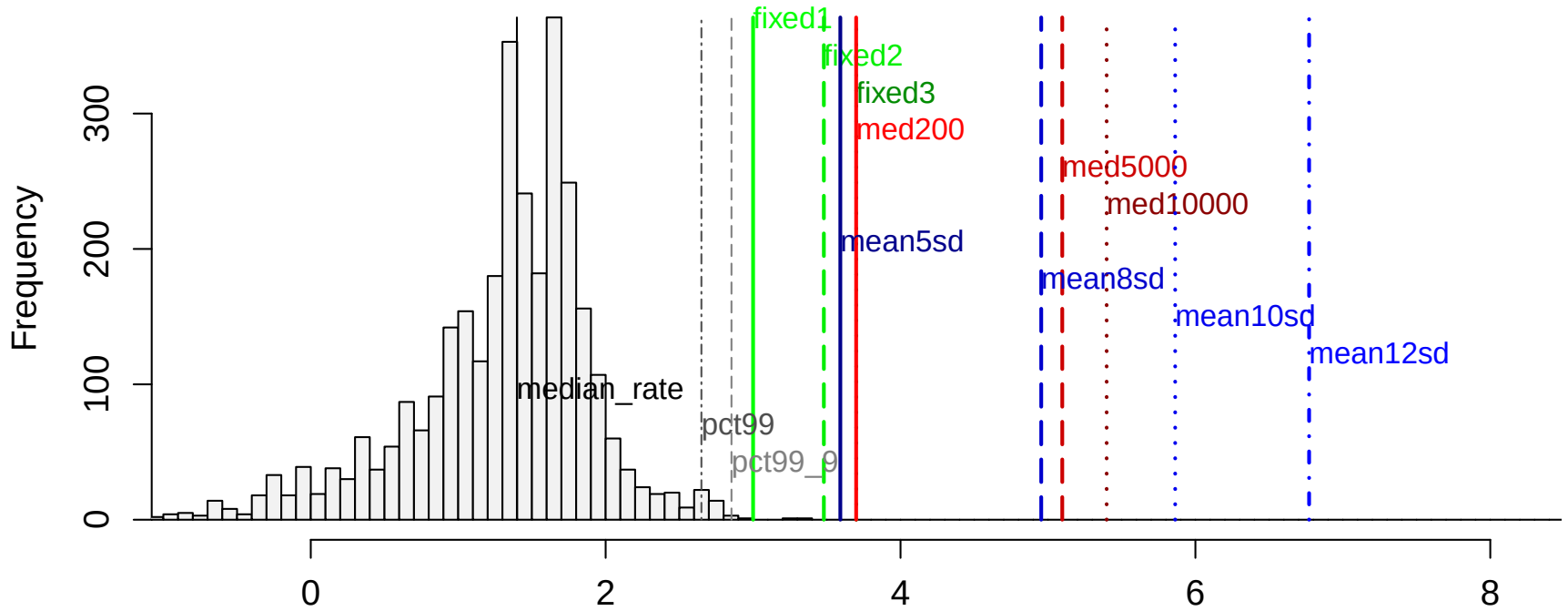
Visual Inspections of Distributions

Sulfuryl fluoride: pounds of AI per application

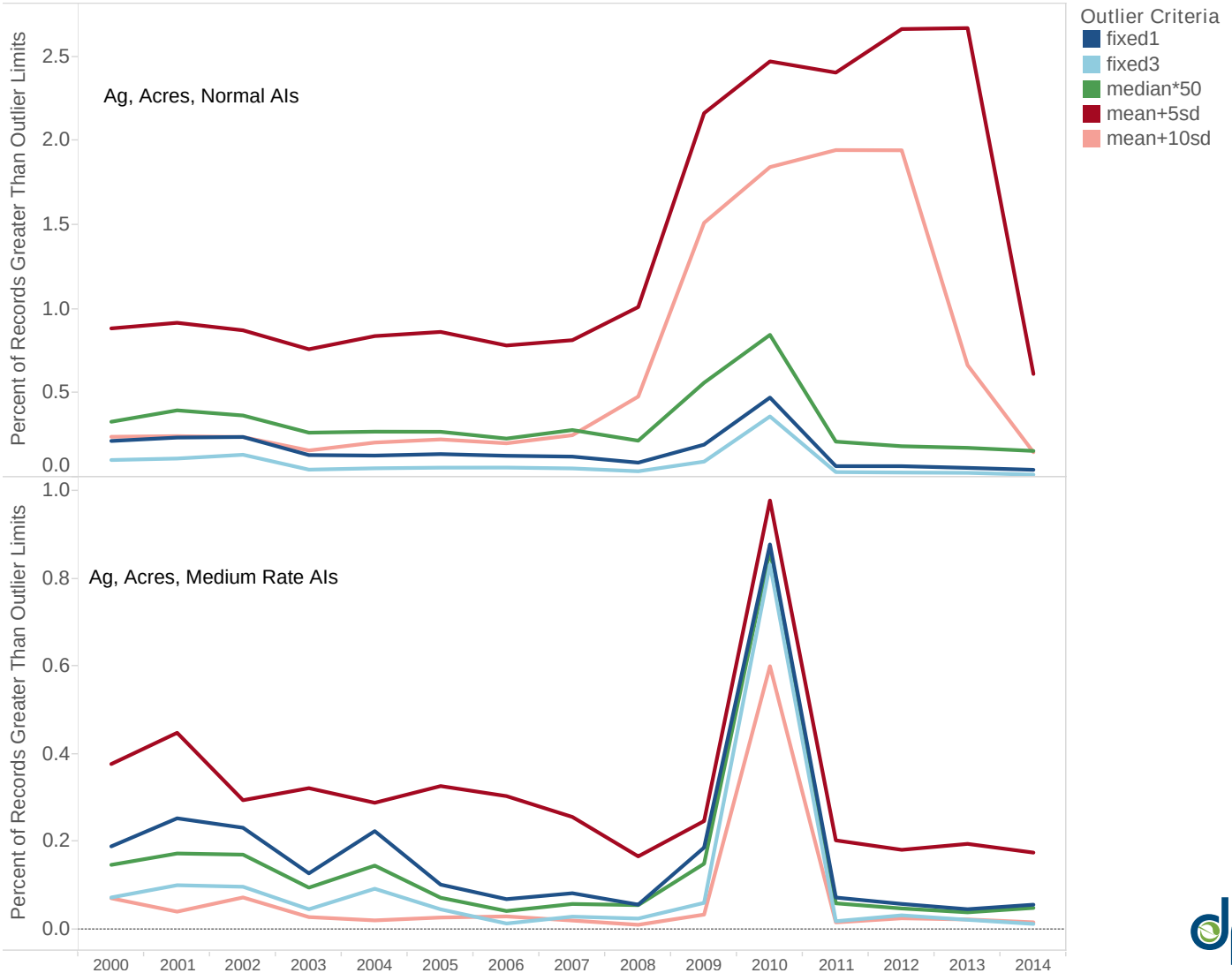


Visual Inspections of Distributions

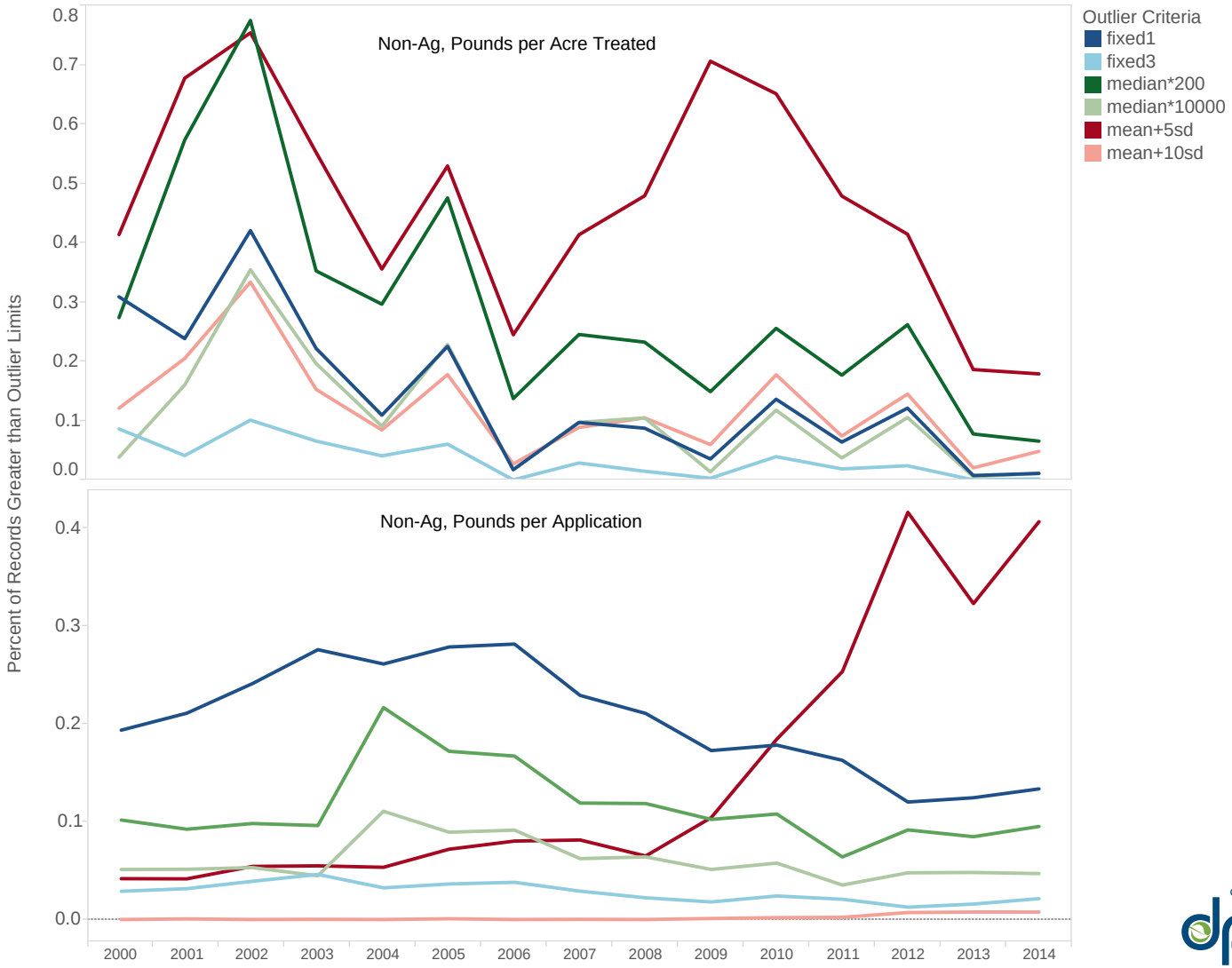
PCNB: pounds of AI per application



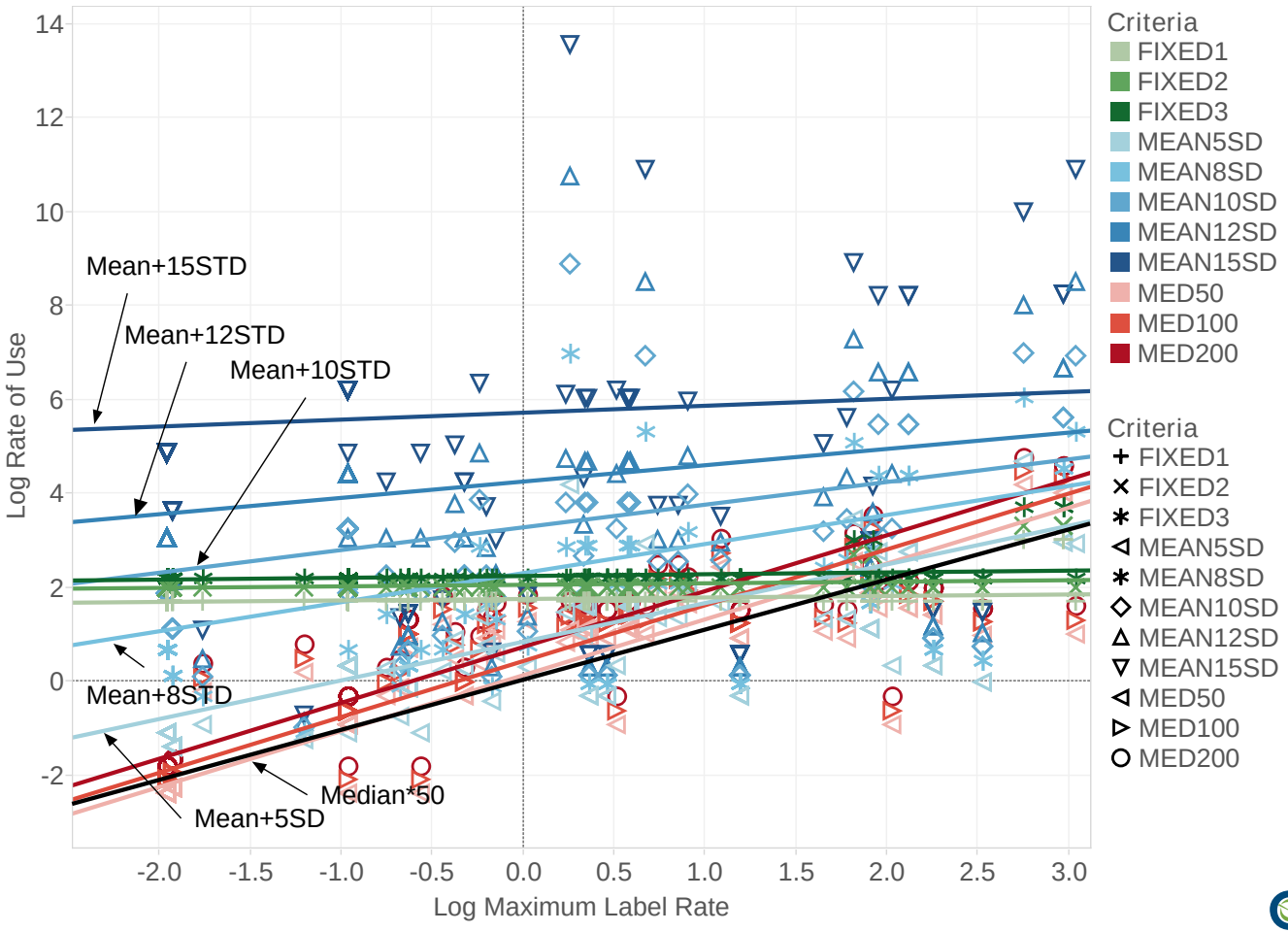
Percent of PUR Records Greater Than Each Limit



Percent of PUR Records Greater Than Each Limit



Applying Criteria to Data With Known Maximum Label Rates



California Environmental Protection Agency
Department of Pesticide Regulation

Software Tools

- Spreadsheets (Excel)
- Database languages (SQL)
- Statistical software (SAS, R)
- Programming languages (R, Python)
- Machine learning (scikit-learn, Weka, TensorFlow)

Excel

- Excel can calculate basic statistics such as subtotals, means, variances, regressions.
- Excel can make basic graphs.
- Excel pivot tables are extremely useful.
- However, Excel cannot load very large files.

Excel

Pivot table of pounds of AI for several AI types and AIs from 2006 to 2010

COUNTY	(All)
MONTH	(All)
PRODUCT_NAME	(All)

Sum of LBS_AI		YEAR				
AI_TYPE	AI	2006	2007	2008	2009	2010
FUMIGANT	1,3-DICHLOROPROPENE	802,009	1,153,900	700,677	595,323	942,279
	METHYL BROMIDE	80,048	96,026	45,124	33,362	14,480
	CHLOROPICRIN	3,301	11,037	34,959	13,340	27,853
	PROPYLENE OXIDE	18,758	11,705	26,683	13,296	14,167
	SODIUM TETRATHIOCARBONATE	4,768	33,941	10,890	25,184	5,859
FUMIGANT Total		908,885	1,306,609	818,333	680,506	1,004,637
FUNGICIDE	ZIRAM	811,613	497,498	295,725	293,405	557,764
	COPPER	800,869	465,229	454,907	282,161	434,652
	IPRODIONE	136,705	117,687	112,086	124,358	193,562
	CHLOROTHALONIL	151,908	100,882	100,944	99,080	217,432
	MANEB	287,224	152,770	85,170	14,173	8,195
FUNGICIDE Total		2,188,320	1,334,065	1,048,833	813,179	1,411,605
HERBICIDE	GLYPHOSATE	1,413,636	1,584,595	1,523,090	1,543,935	1,845,840
	PENDIMETHALIN	163,003	314,614	318,675	359,765	312,197
	PARAQUAT DICHLORIDE	261,950	281,771	230,231	220,351	183,011
	OXYFLUORFEN	215,643	223,901	236,421	226,373	205,954
	ORYZALIN	247,488	194,152	163,735	205,867	199,196
HERBICIDE Total		2,301,720	2,599,033	2,472,151	2,556,291	2,746,198
INSECTICIDE	OIL	12,185,430	10,737,231	12,002,893	11,349,317	11,751,887
	CHLORPYRIFOS	544,946	419,377	290,480	330,593	262,002
	PHOSMET	255,725	189,601	152,350	26,078	22,581
	PROPARGITE	132,644	119,480	108,208	72,040	50,919
	METHOXYFENOZIDE	50,144	55,964	63,103	37,043	53,746
INSECTICIDE Total		13,168,889	11,521,653	12,617,034	11,815,072	12,141,136
Grand Total		18,567,814	16,761,360	16,956,351	15,865,047	17,303,576

SQL

Advantages

- SQL is the most direct way to get specific PUR data.
- You get the most up-to-date data.
- You get quick results.

Disadvantages

- You need to learn a new language.
- It is easy to make mistakes.

SQL

- Simple queries are pretty easy.
- For example, get the total pounds of pesticide product by year:

```
SELECT    year, SUM(lbs_prd_used)
FROM      pur
GROUP BY year;
```

- However, queries can get complicated and it is easy to make mistakes.

SAS

www.sas.com

- SAS is a commonly used statistical program.
- From SAS you can run standard SQL queries.
- The advantage of SAS is that you have statistical procedures not available in SQL.
- SAS can handle large datasets and can read Oracle tables directly.
- However, SAS has an awkward and limited programming language.

R

www.r-project.org

- R is an open source statistical package, that can do most of what SAS can and a lot more.
- R is also powerful programming language.
- R can read Oracle tables directly.
- I started using R because of its excellent graphing capabilities.

Python

python.org

- Python is an open source general purpose programming language.
- Python has many packages for a wide range of purposes.
- One set of packages (numpy and pandas) provides database functionality
- With other packages Python can do much of what you can do in R

Visualization Software

- Many options: recent article on the top 74 visualization software: http://www.predictiveanalyticstoday.com/top-data-visualization-software/
- Tableau [commercial]
- SAS: JMP, Visual Analytics [commercial]
- R: lattice, ggplot2, ggally [open source]
- Python: matplotlib, Seaborn, Bokeh, ggplot (based on R ggplot2) [open source]
- GIS: ArcMap [commercial]; GRASS, QGIS, postGIS [open source]

Visualization software

- WEAVE is an open source, open access, visualization tool suitable for many tasks including business intelligence. http://iweave.com/
- ggobi: GGobi is an open source visualization program for exploring high-dimensional data. http://www.ggobi.org/.
- Mondrian: Mondrian is a general-purpose statistical data-visualization system, for interactive data visualization. http://www.theusrus.de/Mondrian/

Visualization software

- D3.js: is the most versatile JavaScript data visualization library available. https://d3js.org/
- NVD3, C3.js, xCharts, and Dimple offer an abstraction layer on top of d3.js, which makes it easier to draw simple graphs.
- HighCharts: One of the most mature browser-based graphing libraries. See http://shop.highsoft.com/highcharts.html.
- Google Charts: a free charting library with a wide range of graphs. See https://developers.google.com/chart/.
- Inkscape: an open-source vector graphics editor (similar to Adobe Illustrator)

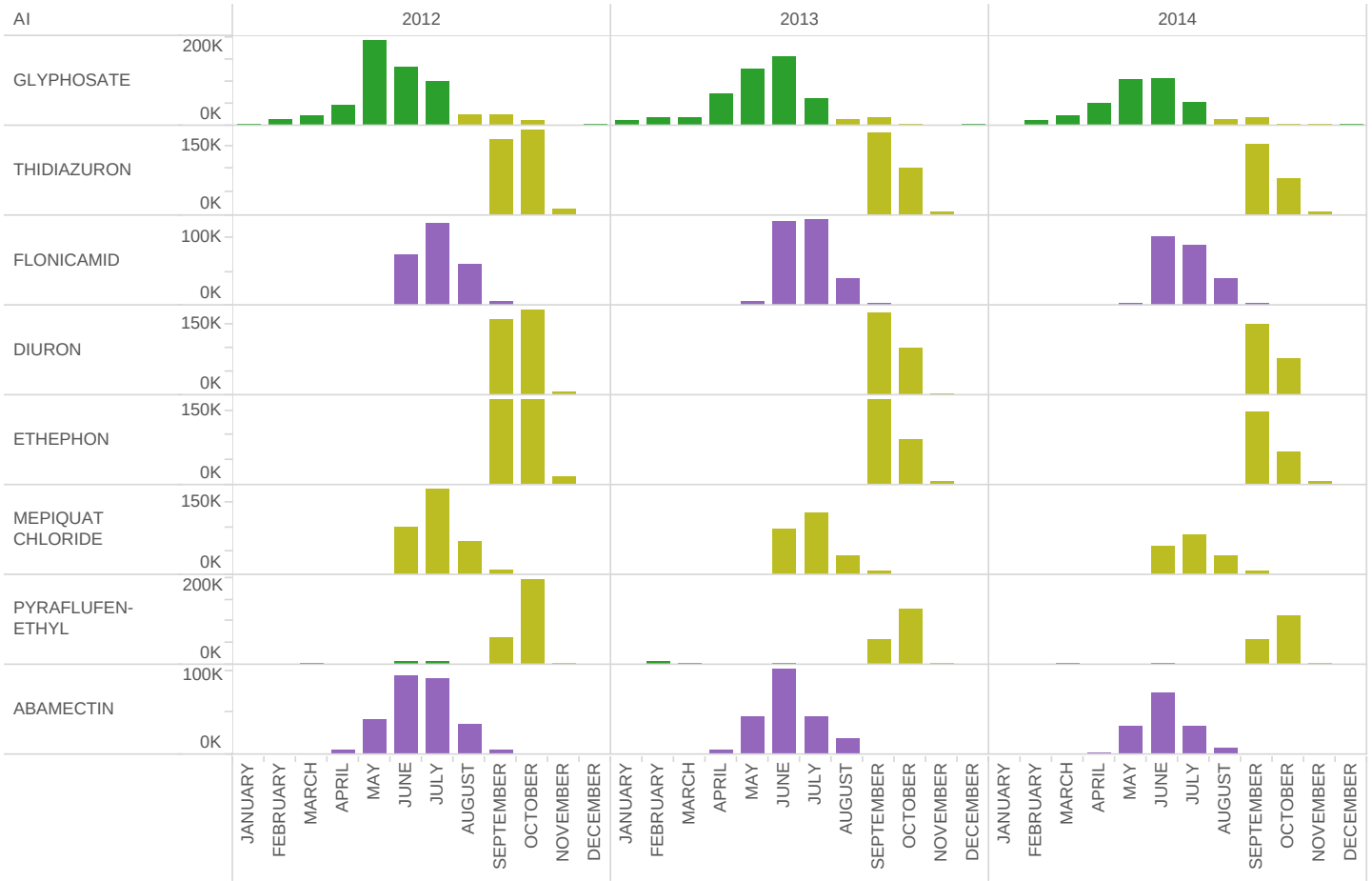
Tableau

www.tableausoftware.com

- Tableau is data visualization program.
- You can directly access large database tables.
- You can create larger and more powerful pivot tables than in Excel.
- You can create many more kinds of graphs.

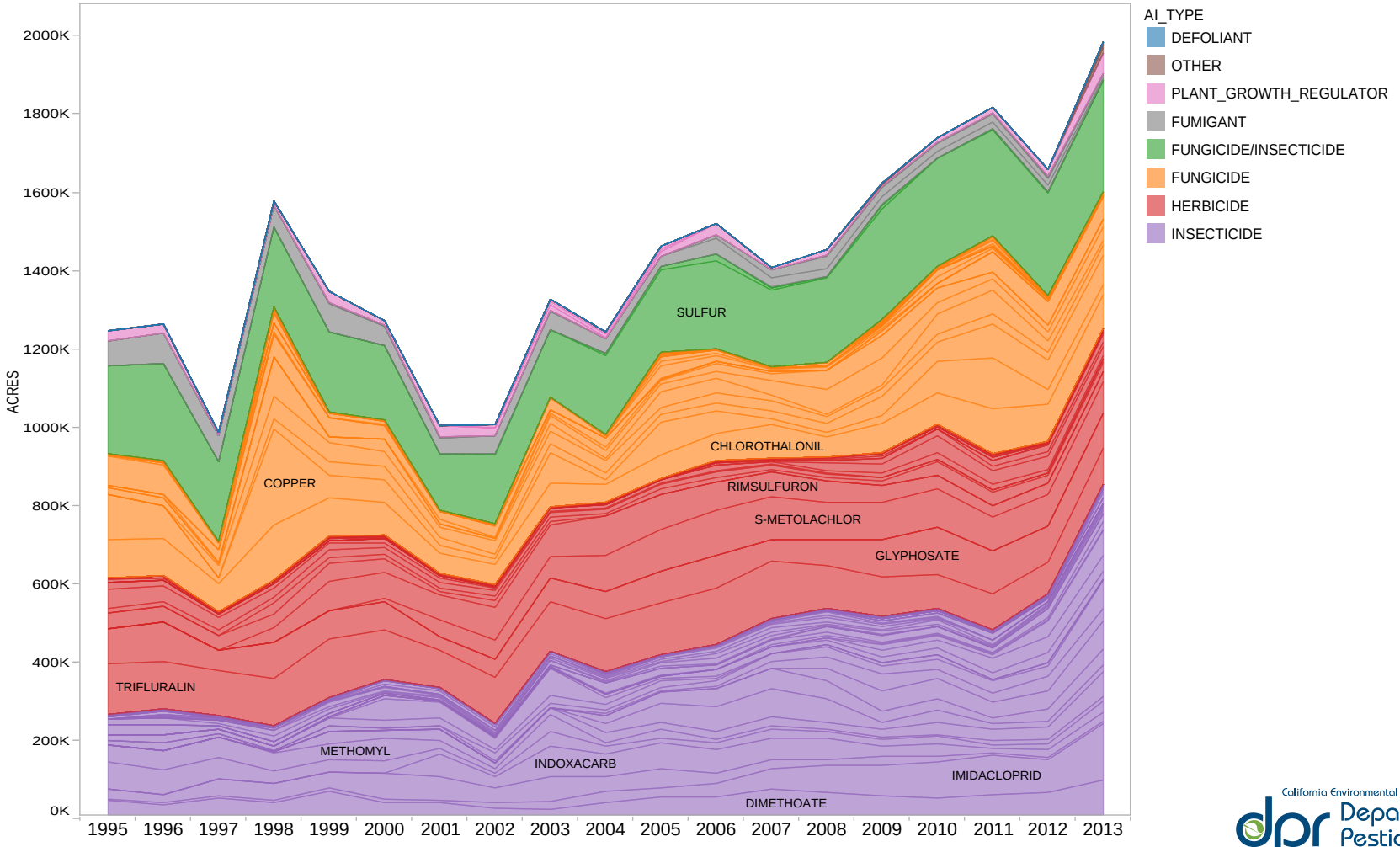
Tableau

Acres Treated in Cotton by AI, AI Type, Year, and Month



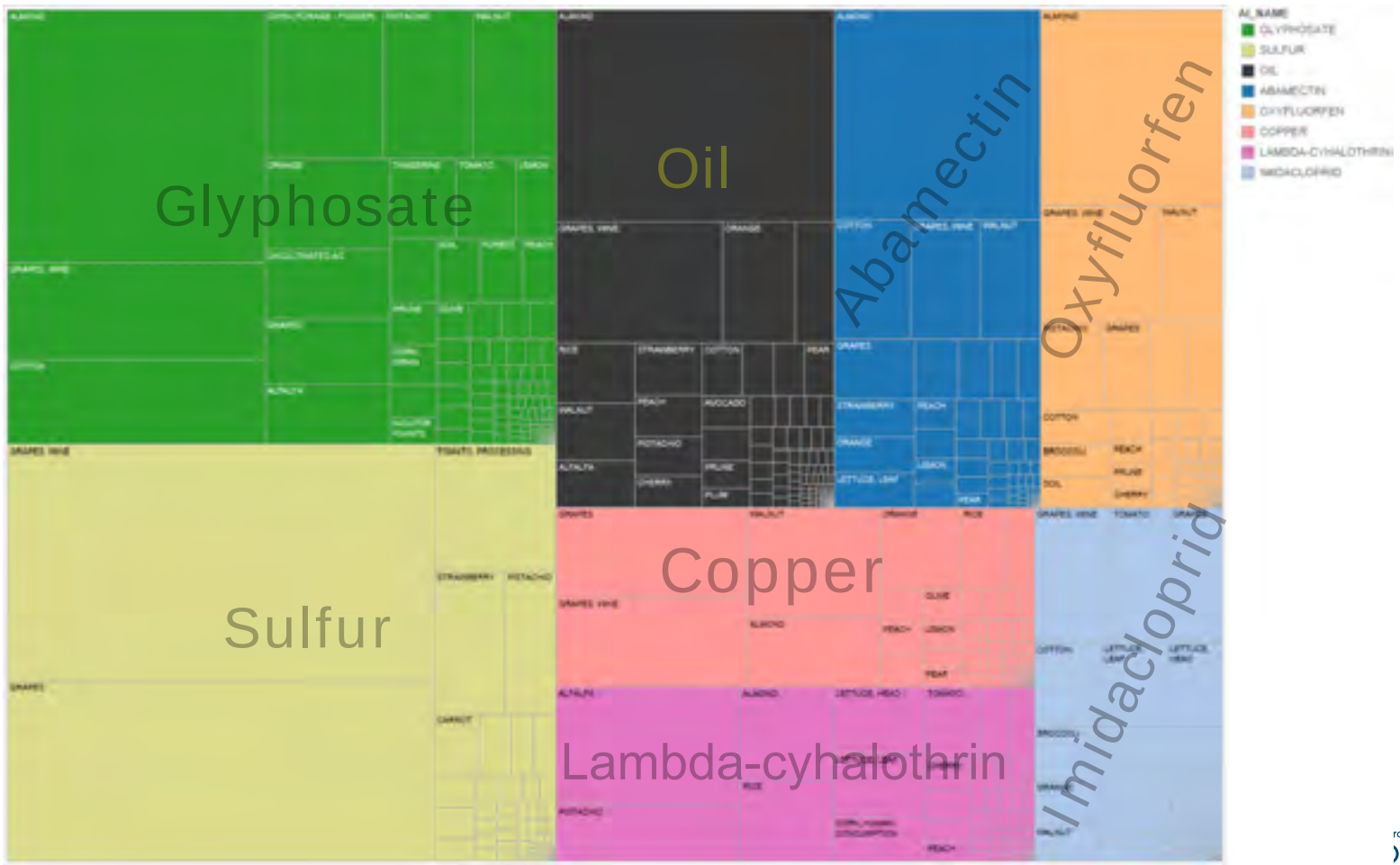
Tableau

Acres Treated in Tomato by AI Type, AI, and Year



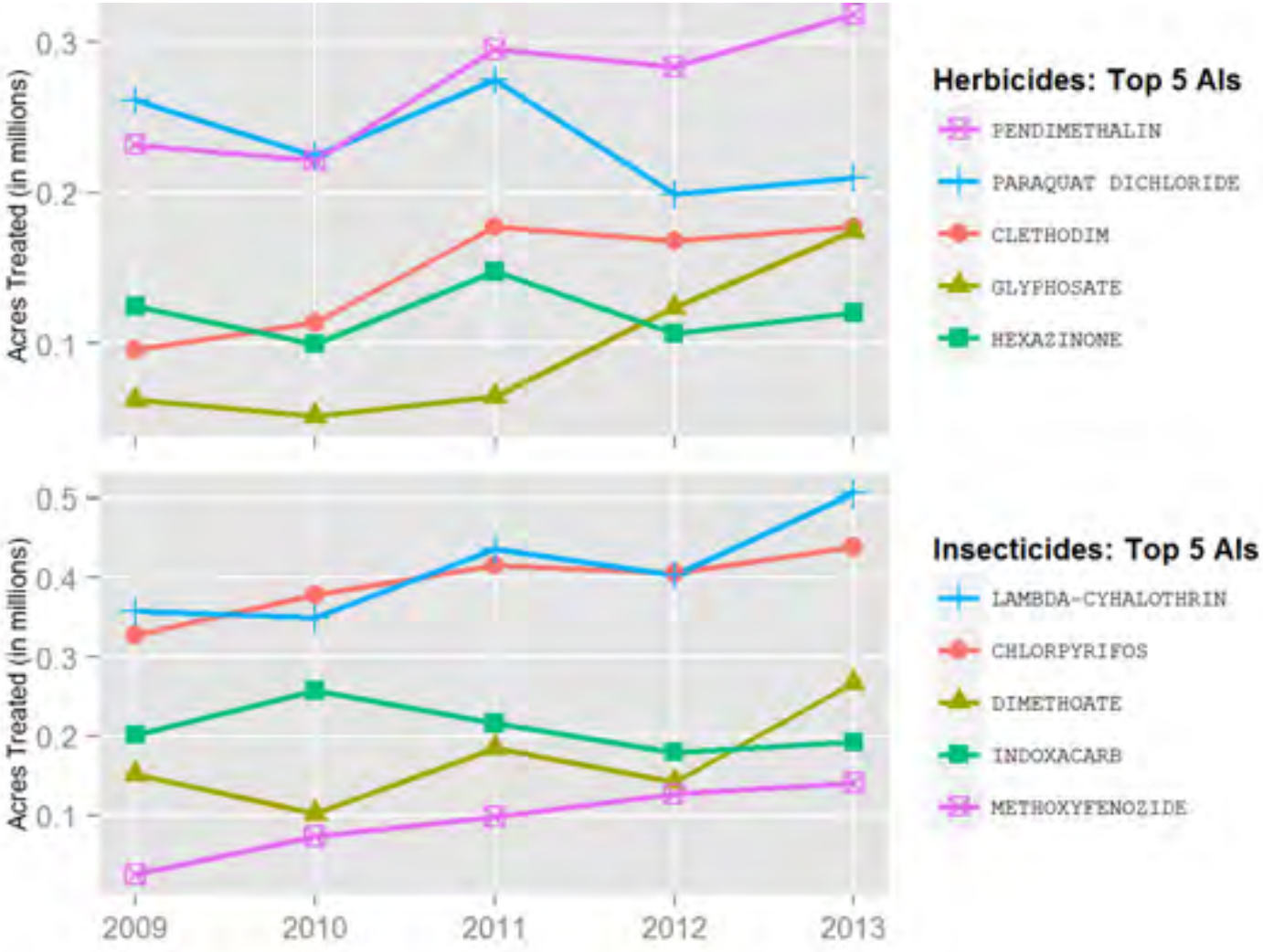
Tableau

Acres Treated by Major AIs on Each Crop in 2013



Graphs created with R

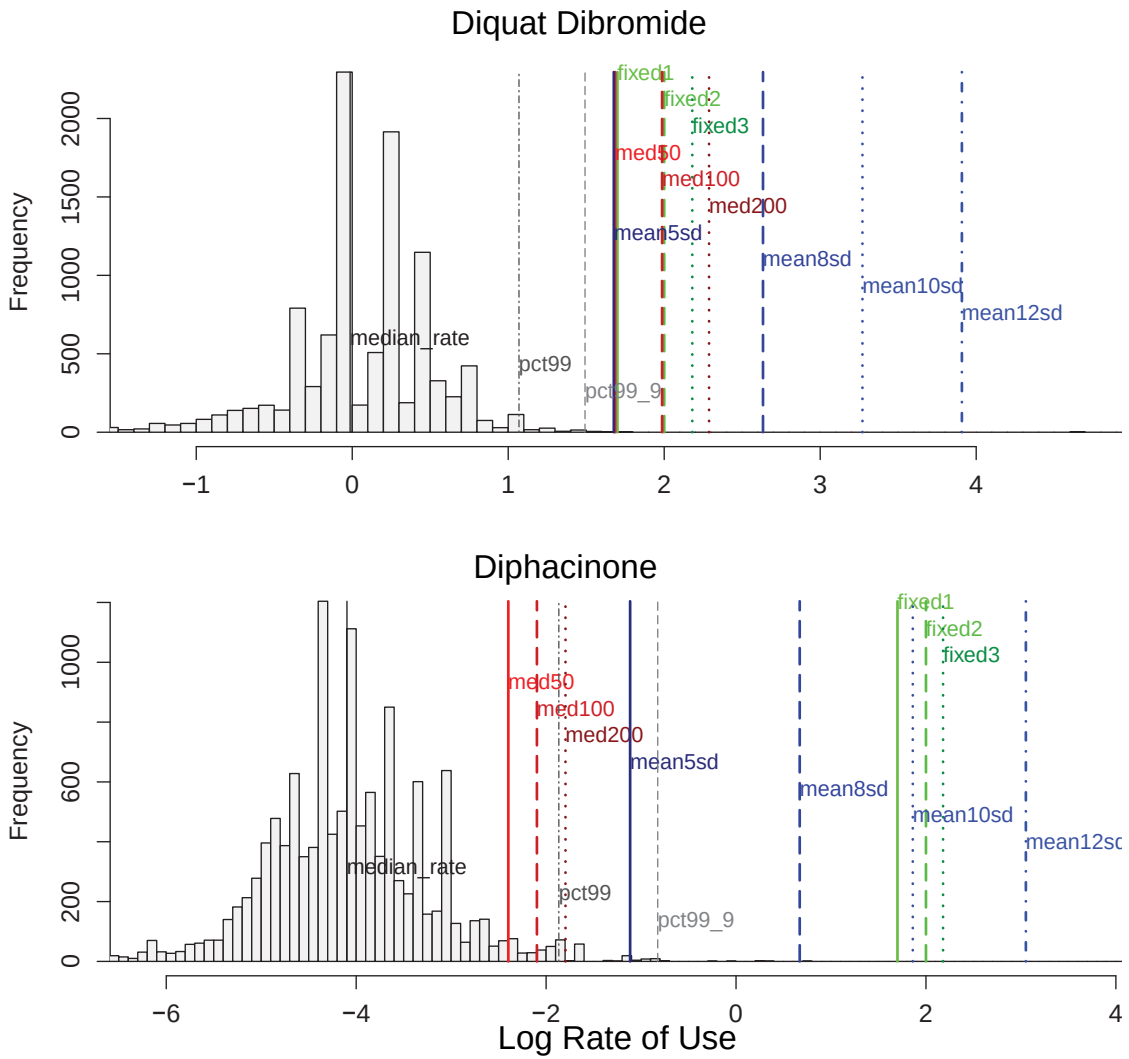
PUR Annual Report: Top AIs by Type in Alfalfa



Protection Agency
rtment of
ide Regulation

Graphs created with R

Frequency Distributions of Rates with Outlier Limits



Questions?