Procedures for Identifying High Rates of Pesticide Use

Larry Wilhoit

September 30, 2016

Contents

1 Introduction **1**

2 Detecting outliers in rates of pesticide use **2**

3 Current criteria for identifying outliers **2**
3.1 Criteria definitions . . . 2
3.2 Advantages, problems, and improvements for each criterion . . . 3
3.2.1 Fixed limits . . . 3
3.2.2 Median rate × 50 . . . 3
3.2.3 Median + MAD × 50 . . . 4
3.2.4 Neural network . . . 4

4 Distributions in rates of use **5**
4.1 Typical distributions . . . 5
4.2 Multimodal distributions . . . 6
4.3 Distributions with high variance . . . 7
4.4 Distributions with low variance . . . 7
4.5 Distributions in pounds of AI per application . . . 7

5 New criteria for identifying outliers **10**
5.1 Pesticide label rates . . . 11
5.2 Fixed limits . . . 11
5.3 Median rates times a factor . . . 17
5.4 Trimmed mean + standard deviation times a factor . . . 18
5.4.1 Defining AI groups . . . 18
5.4.2 Results from the procedure to find AI groups . . . 20

6 Criteria evaluation **25**
6.1 Visual inspections of distributions with the different outlier limits . . . 25
6.2 Number of records identified as outliers by each criterion . . . 31
6.2.1 Evaluation of fixed limits . . . 31
6.2.2 Evaluation of all criteria . . . 39
6.3 Relative ranking of outlier limits . . . 41
6.4 Applying criteria to data with known maximum label rates . . . 44

7 Conclusions **46**

Bibliography **50**

1 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. The PUR contains information on nearly all production agricultural pesticide use and on some non-agricultural use in California, amounting to an average of 2.6 million records each year from 1990 to the present time, and this number has increased in the last few years. The data collected includes the pesticide product used, the amount of pesticide applied in both pounds and the unit used in the original report (such as gallons or ounces), the area or other unit treated, the site or crop treated, the geographic location to a square-mile section, the date it was applied, the grower and field identifier, and the application method. Although only the pesticide product is reported, the pesticide active ingredients (AIs) can be found by linking the PUR with another DPR database that lists all AIs in each product and the percent by weight of each AI in the product.

The PUR contains two types of records: 1) production agricultural use, which includes applications to growing crops or agricultural fields; these uses will be referred to as "agricultural uses"; and 2) all other kinds of applications, which includes post-harvest commodity treatments, rights of ways, landscapes, structural use, and other non-agricultural uses by commercial applicators; for convenience these will be referred to as "non-agricultural use" even though some of these uses might be thought of as agricultural. In the database the type of record is identified by the field record_id. Record_id values of "A", "B", "E", "F", "1", or "4" are production agricultural records; values of "C", "G", or "2" are non-agricultural records. Each production agricultural record represents a single application of a particular pesticide product on a field on a given day. These records always contain the area or other unit treated. The non-agricultural records are a sum of all applications of a particular pesticide product by an applicator to a particular site in a county for an entire month. These records sometimes contain a unit treated and sometimes not, but they usually report the total number of applications. If a non-agricultural record reports the unit treated, then the rate of use can be calculated. Even if no unit treated is reported another kind of rate can be calculated, pounds of per application, assuming there is 1 application if none is reported. However, because of inconsistencies in reporting, these two kinds of rate for non-agricultural records are less reliable than for agricultural records.

In the PUR units treated are reported as either acres, square feet, cubic feet, 1000 cubic feet, pounds of material treated, tons treated, or a miscellaneous unit (such as plant, bins, or holes). Some of the units are comparable and can be converted into the same units. For all analyses done in this report, square feet was converted to acres, 1000 cubic feet was converted to cubic feet, and tons were converted to pounds. When comparing rates, the units treated must be the same. For example, pounds per acre is a completely different measure than pounds per cubic feet or pounds per pound of material treated.

After a pesticide application is made, the grower or some other designated person submits a record of this application to their County Agricultural Commissioner's (CAC) office. Some of these reports are on paper forms, but more people are submitting the reports online. These reports are entered into a county pesticide use database; if submitted on a paper form, county staff manually enter the data into the database. These data are then submitted electronically to DPR's statewide PUR database.

Because of the importance of the PUR for many groups and individuals, the database needs to be as accurate and complete as possible. The data are screened for many possible errors at data entry, at the CAC offices, and at DPR before they are entered into the PUR database. One of the most important errors occurs in the amount used. It is easy to misplace a decimal or to report the wrong unit (such as gallons rather than ounces). This kind of error can often be found by looking at the rate of use (pounds of pesticide per unit treated or pounds per application). Currently, the error checking procedures only look for unusually high rates, but not unusually low rates, because it is the

high values that can most seriously affect an analysis. However, the current procedures for this check have a number of deficiencies. These procedures were developed nearly 20 years ago, and since then, information has been collected about problems with the procedures, and new methods have been developed for fixing most of these problems. There is also more PUR data that can be used to test different procedures.

This document will present improved methods for identifying outliers in rates of use. Different groups of people that use the PUR have developed their own methods for identifying high rates of use. DPR would like these groups to review and discuss these proposed procedures and hopefully come to an agreement on the best ones to adopt. Having procedures that everyone accepts would allow different analyses to use the same data and would reduce the need for different people to spend time developing their own procedures. More accurately identifying errors will save time for people doing analyses and provide them greater confidence that the data are as accurate as possible. Reliable procedures will also save CAC staff time. When outliers are incorrectly identified, county staff waste time in trying to determine corrections for data that are, in fact, valid.

2 Detecting outliers in rates of pesticide use

The term "outlier" does not have a formal statistical definition; it is just an unusually extreme value. What is "unusual" is partly a subjective judgment. However, a number of statistical procedures have been developed to detect outliers, such as Rosner's, Dixon's, and Grubbs' test (Barnett and Lewis, 1994). Nearly all of these assume a specific probability distribution of the data, usually a normal (or Gaussian) distribution. Since the distributions of rates of use can be highly unusual and variable, there can be no clear, objective statistical tests for rate outliers. Nevertheless, some kind of automated method needs to be found that can flag likely outliers in the PUR. Any method will have to consider what pesticide is used since rates of use vary tremendously from pesticide to pesticide. The median rate of use varies between that of chlorophacinone at 0.00002 pounds per acre to that of metam-sodium at 207 pounds per acre, a 10 million fold-difference. Rates often also depend on the crop or site treated or the pest being treated.

3 Current criteria for identifying outliers

3.1 Criteria definitions

For most of the past 18 years, the following four basic criteria have been used to identify possible outliers in rates of use (Wilhoit, 1998):

1. Fixed limits. Records are flagged if the pounds per acre of any non-fumigant active ingredient is greater than 200 or if the pounds per acre of a fumigant is greater than 1000. The fumigants include methyl bromide, chloropicrin, 1,3-dichloropropene, metam-sodium, metam-potassium, and dazomet.

2. Median rate $\times$ 50. Records are flagged if the pounds of a pesticide product per unit treated is greater than 50 times the median rate for all uses of that product on the same site, the same unit treated, and the same record type during the most recent PUR year. There are two record types: applications to production agricultural fields and all others, mostly non-agricultural applications.

3. Median + MAD × 50. Records are flagged if the pounds of a pesticide product per unit treated is greater than the median plus 50 times the median absolute deviation (MAD). As with the previous criterion the median is calculated from the set of all rates of the same use type.

4. Neural network. Records are flagged if the pounds of a pesticide product per unit treated is greater than a limit value calculated using a neural network procedure.

The first two criteria flag the most extreme rates of use. The third criterion had serious problems and was not generally used. The neural network flagged more records than the first two, but for technical reasons has not been used since 2010. Ideally, all records flagged by these three useful criteria should be checked for accuracy, but there are simply too many records to do so. Therefore, only the most serious outliers are reviewed; specifically, records flagged by either of the first two limits are sent back to the county for review by their staff. However, the PUR contains flags on all records identified by each of these criteria so a person analyzing the data can do their own research on the validity of flagged records.

3.2 Advantages, problems, and improvements for each criterion

3.2.1 Fixed limits

The advantage with the fixed limit criterion is that it does not depend on previously reported uses. Thus it could be useful for situations where there were very few or no previously reported uses, such as with new AIs, or where most of the previous records had errors. For example, if one of two previously reported uses had an erroneously high rate, that rate or later reported high rates would not be flagged by the other criteria. But a fixed limit could identify this as a high rate.

One problem with this criterion currently is that it is used only for records with units in acres. That problem is easily remedied by determining fixed limits for other units as well. Another problem is that several non-fumigant pesticides are used at high rates, though their rates are not usually as high as most fumigants. The non-fumigant fixed limit (200 pounds per acre treated) is too high for most AIs. An improvement would be to classify all AIs into three or more groups based on their typical rates of uses. Then more accurate fixed limits could be set for all AIs.

3.2.2 Median rate × 50

The advantage of the second criterion (median × 50) over the fixed limit criterion is that the outlier limit is tuned for each pesticide and is most useful for pesticides used at low rates. Because the fixed limit is set high relative to the typical uses of most AIs, it will often miss errors with pesticides that are used at very low rates.

One problem with the second criterion is that it does not account for the dispersion or spread in rates of use. For pesticides that are used with widely differing rates, some of the valid high rates would be incorrectly flagged as outliers by this criterion. For pesticides applied with a very narrow range of rates, many incorrect high rates would not be flagged. Another problem is that it is often based on few records, which occurs when there are few previously reported uses during the most recent year for that pesticide and use type.

An improvement for this would be to base the criterion on uses of the same AI, regardless of product or site, thus providing more records for determining unusually high rates. It could also be improved by using rates from several years rather than just one year. Finally, combining some of the units

treated would increase the number of records. For example, there are two units for measuring areas — acres and square feet — and two measures of volume — cubic feet and 1000 cubic feet. Square feet can be converted to acres, and 1000 cubic feet to cubic feet, thus giving more records for each unit.

The effect of using all rates for an AI, rather than each product and site, is seen by comparing the number of PUR records for each product, site, unit treated, and record type with the number of records for each AI, unit treated, and record type. There are 90,120 combinations of product, site, unit treated, and record type in the PUR from 2010 to 2014, and the median number of records for each combination is 6, which is a very small sample to use in creating a frequency distribution. For agricultural records with units in acres, the number of combinations of product and site is 73,110 and the median number of records is 7.

By contrast there are 2,625 combinations of AI, combined unit treated, and record type, and the median number of records per each combination is 20, which may be sufficient for getting accurate distributions. The difference is more pronounced for agricultural records with units in acres, where the number of AIs is 1,026 and the median number of records is 635. Thus, using sets of PUR records with AI, unit treated, and record type will provide more accurate frequency distributions for more PUR records and, thus, provide better determinations of outliers.

Of course, the disadvantage of using AIs rather than products is that application rates for some AIs differ for different products and sites. However, rates of use are considered outliers only if they differ by nearly 2 orders of magnitude, which is much larger than most differences between different products or sites of one AI. There are a few cases, however, where rates on particular products or sites are very different, so some method is needed to handle these situations (see section 5.4.1, page 18).

3.2.3 Median + MAD × 50

The third criterion, like the second criterion (median × 50), is also based on sets of PUR records with the same product, site treated, unit treated, and record type, and thus has the some of the same advantages and disadvantages as the second criterion. Similarly, it can be improved by using records with the same AI, unit treated, and record type. However, the third criterion has the additional advantage that it considers the dispersion of rates, which was ignored in the second criterion. Dispersion can be measured by the variance or a robust measure of variance, such as the MAD, which was the measure used in the current criterion. This criterion would seem to hold the most promise and is in fact the basis of one of the new proposed criterion. However, as currently implemented there are so many situations where it incorrectly identifies outliers, that it was nearly useless. The problems appeared because of the unusual kinds of distributions that occur in pesticide rates. The problems and possible solutions will be explained in section 5 of this report.

3.2.4 Neural network

The final current criterion uses neural networks (Wilhoit, 1998). This technique was used because previous attempts to find a criterion based on statistical measures failed. However, now with more data and new ideas, this method is no longer needed. The disadvantages with a neural network are that it is more difficult to implement and it operates as a black box providing little understanding of how the outlier limits are determined.

4 Distributions in rates of use

4.1 Typical distributions

To see some of the problems and possible solutions to identifying outliers in rates of use, several example distributions will be examined. The frequency distribution of rates of use for glyphosate in 2014 (Figure 1) is typical of many AIs. The problem with using this graph to identify outliers is obvious — because of a large outlier, at 17,200 pounds per acre, all the other rates of use are grouped together in one histogram bar, and no information is provided for the distribution of these rates. The distribution of more typical rate values, less than 15 pounds per acre, is more useful for finding outliers (Figure 2). This kind of distribution is typical, with many uses around 1 pound per acre and fewer uses at higher rates, giving a highly skewed distribution. This graph can be improved by using the logarithms, or logs, of the rates (Figure 3). The advantage of using logs is that both low and high rates can be displayed in one graph. Further, this transformation usually creates a distribution closer to a normal distribution, which has better statistical properties. Using logs solves many of the distribution problems, but not all. There are other situations that need to be dealt with, such as multimodal distributions and distributions with low or high variance.

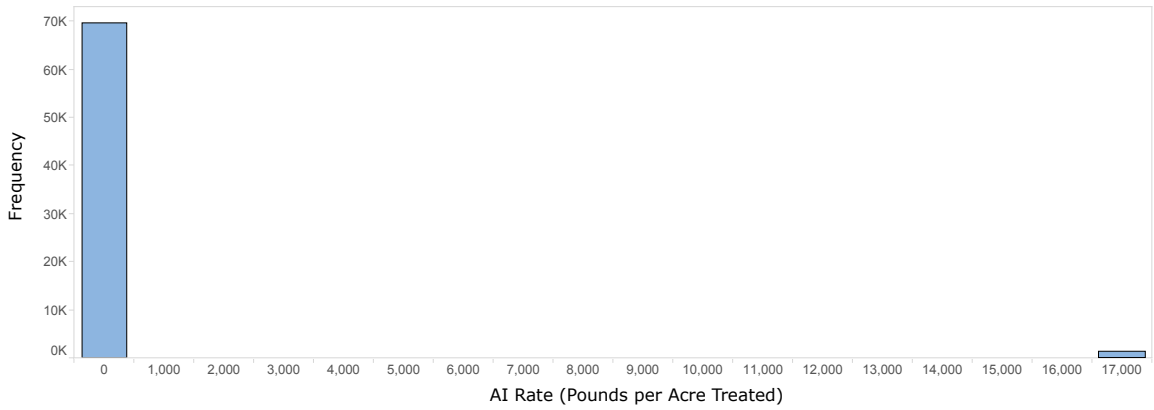

Figure 1: *Frequency distribution of AI rates of use with glyphosate.*

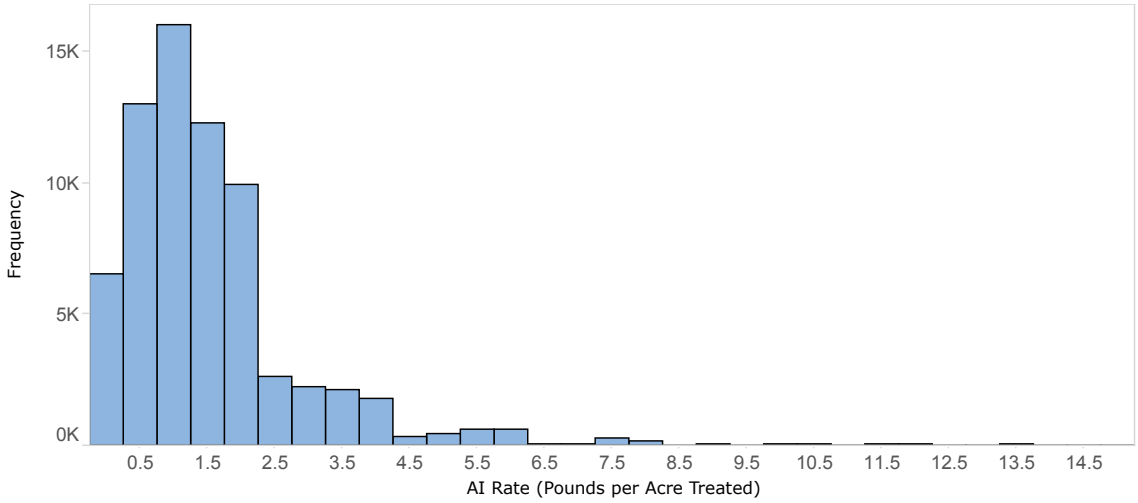

Figure 2: *Frequency distribution of AI rates of use with glyphosate.*

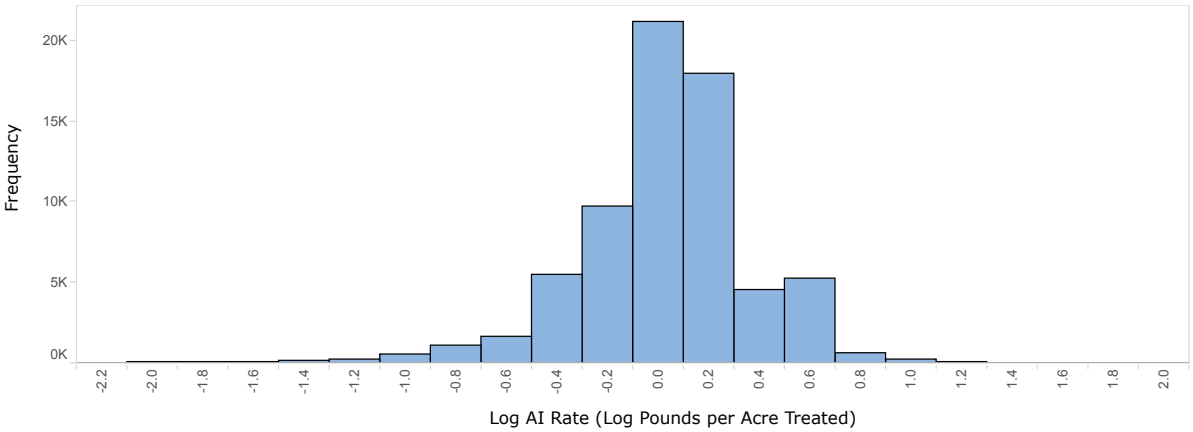

Figure 3: *Frequency distribution of log of AI rates of use with glyphosate.*

4.2 Multimodal distributions

Some distributions have two or more distinct peaks or modes, usually because of distinct use rates on different kinds of crops or from different kinds of pesticide products. One example is the distribution of spinosad rates, which has two distinct peaks (Figure 4). The left, smaller peak comes from bait products, and the other peak comes from all other pesticide products.

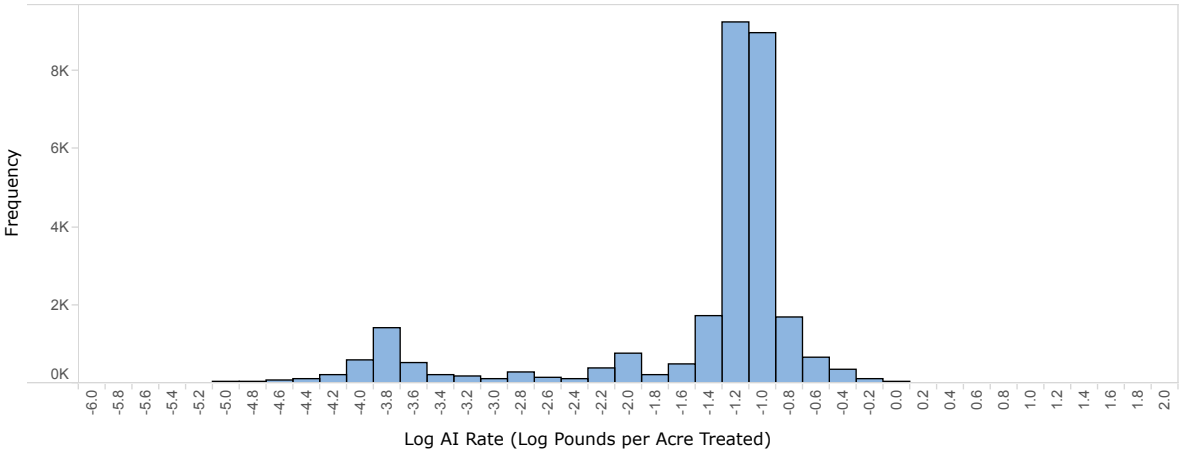

Figure 4: *Frequency distribution of log of AI rates of use with spinosad.*

Another example is the distribution for copper sulfate (pentahydrate), which has two peaks near each other. The lower rate peak is mostly from applications to nurseries and the higher peak is from applications to other kinds of crops. These two situations can be handled by creating subsets of uses. That is, measures of outlier statistics can be based not on all uses of an AI, but on appropriate subsets of similar uses. A method for determining AI groups is explained in section 5.4.1 (page 18).

Some AIs show distributions with more than two modes, such as *Bacillus thuringiensis* subspecies *israelensis* (serotype H-14) (Figure 5). In this case, the different modes cannot be easily distinguished by site treated or product used.

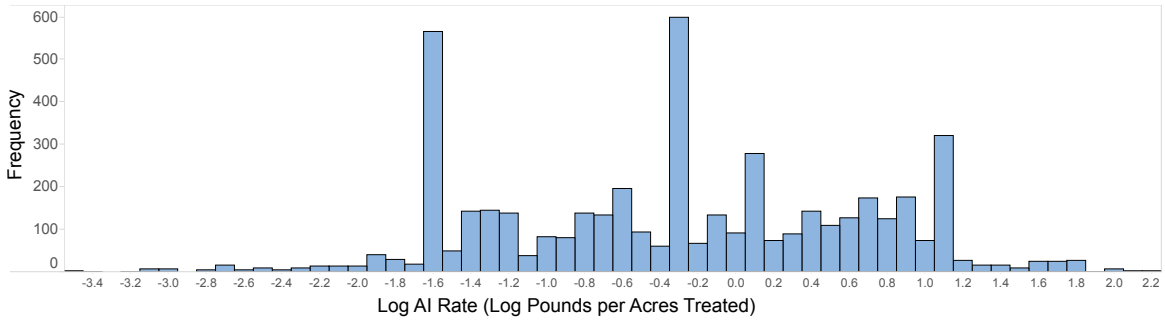

Figure 5: *Frequency distribution of log of AI rates of use with* Bacillus thuringiensis *subspecies* israelensis *(serotype H-14).*

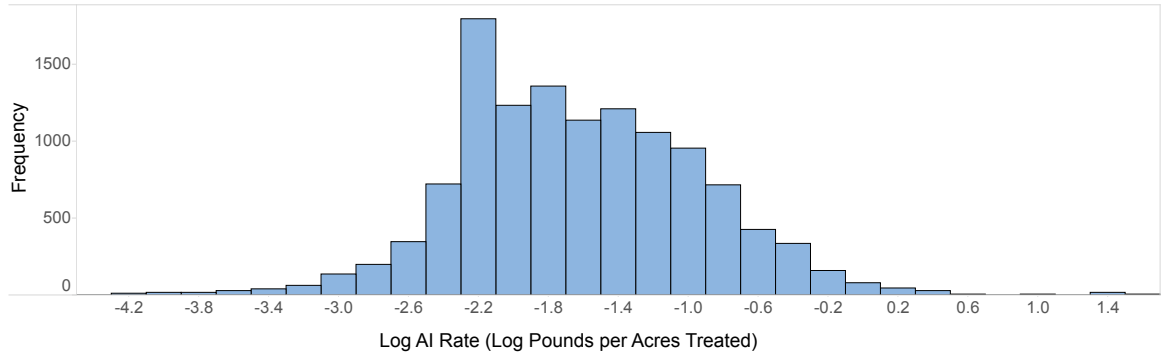

Figure 6: *Frequency distribution of log of AI rates of use with aluminum phosphide.*

4.3 Distributions with high variance

Some distributions of rates have a high variance, that is, the distributions are wide. Such distributions are sometimes a result of several kind of uses, although we may not be able to identify what distinguishes those uses, such as in the previous example, *Bacillus thuringiensis* subspecies *israelensis* (serotype H-14) (Figure 5). Sometimes there just appears to be a wide range of rates, as with aluminum phosphide (Figure 6).

4.4 Distributions with low variance

Metaflumizone presents the opposite situation, with a very narrow dispersion (Figure 7). This situation is actually fairly common. It was for these situations that the current criterion using dispersion failed most seriously. The reason is that if more than half of the records have rates equal to the median then the MAD (the measure of dispersion used in the current criterion) is 0. In that case, the outlier limit will equal the median and will thus flag most of the rates as outliers, which obviously is not correct.

4.5 Distributions in pounds of AI per application

The pounds of AI per application for the non-agricultural applications generally have distributions that are close to a normal distribution. The variances of these distributions are generally larger than that for pounds per acre, which is not surprising since the number of applications are not consistently

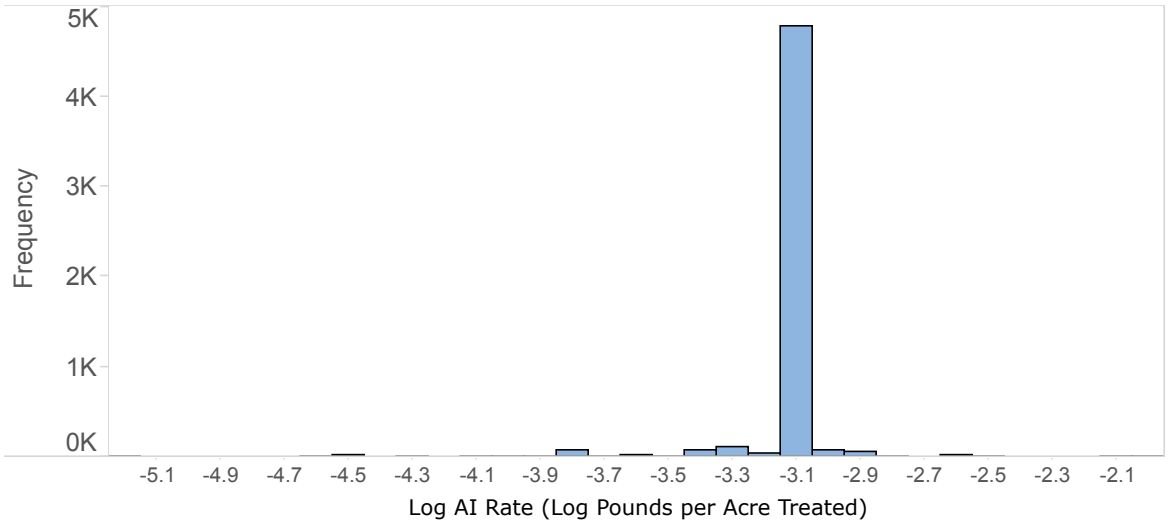

Figure 7: *Frequency distribution of log of AI rates of use with metaflumizone.*

reported, and what counts as one application can vary, for example, from a spot treatment for ants to a treatment of an entire building. An AI with one of the smallest variances in pounds per application is fluazinam (Figure 8). Sulfuryl fluoride has a larger variance, but the distribution still looks close to a normal distribution and illustrates a fairly typical non-agricultural distribution (Figure 9). Some distributions have two modes, such as that for fipronil (Figure 10), and a few AIs have a broader distribution and may represent several distinct kinds of uses, such as dichlobenil (Figure 11).

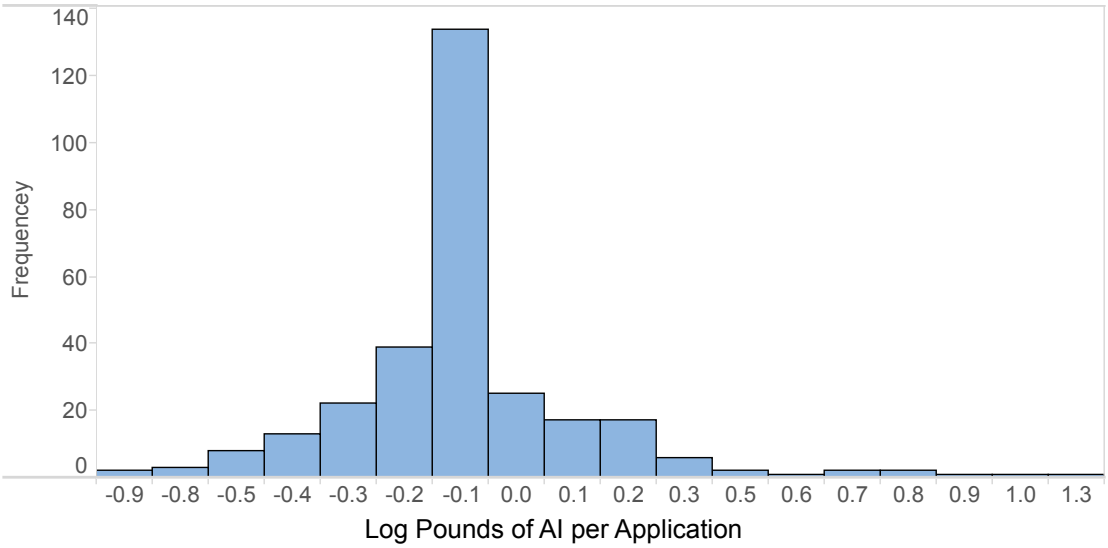

Figure 8: *Frequency distribution of log of AI rates of use with fluazinam.*

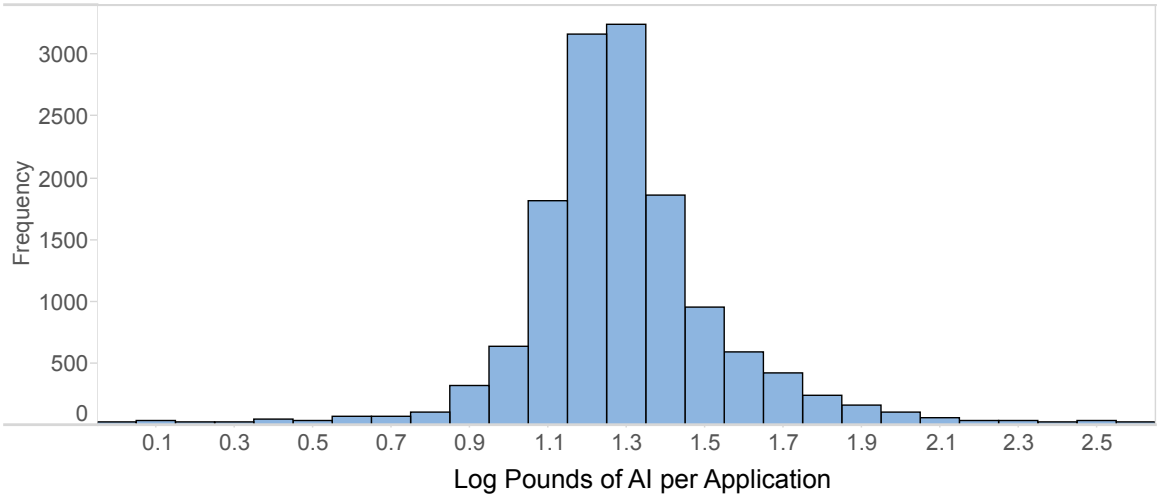

Figure 9: *Frequency distribution of log of AI rates of use with sulfuryl fluoride.*

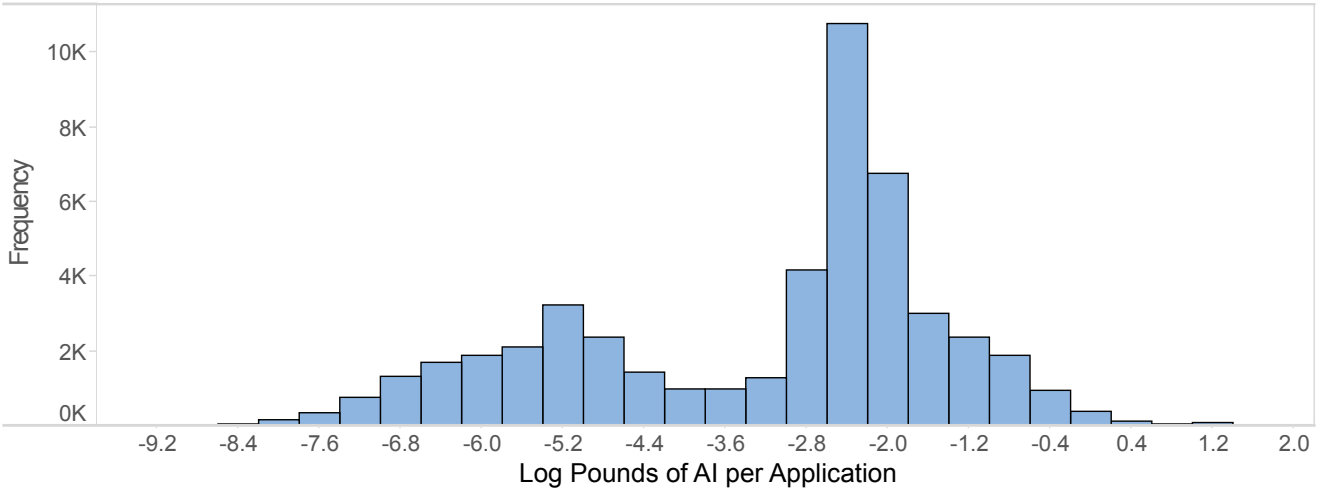

Figure 10: *Frequency distribution of log of AI rates of use with fipronil.*

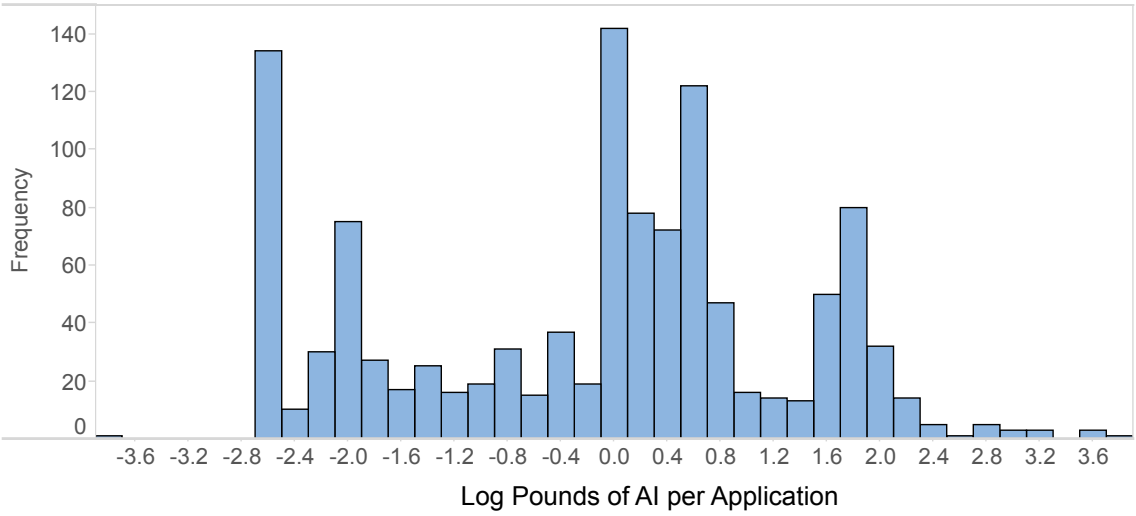

Figure 11: *Frequency distribution of log of AI rates of use with dichlobenil.*

5 New criteria for identifying outliers

From the discussion of the problems with the current outlier procedures, several ideas arose for making improvements:

1. For the fixed limits, develop limits for all treated units, not just for acres.
2. Also for the fixed limits, create more AI groups than just fumigants and all others.
3. For the other limits that are based on previous records, develop limits for each AI, unit treated, and record type rather than for each pesticide product, site treated, unit treated, and record type. This results in larger sets of records to use for determining outliers.
4. Larger sets of records can also be created by using data from more than one year.
5. Use the log transform of rates to create distributions that are closer to a normal distribution.
6. Develop better measures of dispersion to replace the current criterion using MAD.
7. Develop limits for all non-agricultural records. Most of these records do not have values for unit treated, so pounds per unit treated are not available. However, outliers for a different kind of rate, pounds of AI per application, can be examined. The term "rate" used in this report refers both pounds of AI per unit treated and pounds of AI per application.

Outliers in data sets with non-normal distributions have traditionally been identified using statistically robust measures of central location and dispersion. Robust measures of location include the median, trimmed mean, and Winsorized mean; robust measures of dispersion include MAD, Gini's mean difference, Winsorized and trimmed standard deviation, and interquartile range (the difference between the third and first quartile) (Hoaglin et. al., 1983; Wilcox, 2010). These measures provide many different choices each with many different parameters. Several of these measures were studied previously, and from these studies the trimmed mean and standard deviation were determined to provide better methods for identifying outliers. This report discusses the following criteria (the abbreviations that will be used to refer to these criteria in this report are given in square brackets):

1. Pesticide label rates
2. Fixed limits with different limits for different sets of AIs [fixed1, fixed2, fixed3]
3. Median log(rate $\times$ x) [median*x], calculated from previous use records, and where x can be one of several different values to adjust the level of outlierness.
4. Trimmed mean(log rate) + y $\times$ standard deviation(log rate) [mean+y*SD], where y can be one of several different values. This criterion removes the highest and lowest 5% from the set of log rates and then calculates the means and standard deviations from that reduced set.

This list includes several kinds of criteria with different levels of outlierness for each kind. The goal is to determine which criteria and levels work best. More than one criteria or level can be used to flag outliers in the PUR, as is done currently. These criteria will all be evaluated in section 6.

Limits are determined separately for different types of use, which are defined by four factors: by type of record (agricultural or non-agricultural uses), by unit treated (acres, cubic feet, pounds, or miscellaneous unit), AI rate type, and, for the trimmed mean criteria, by a more specific type of use that depends on the AI, product used, or site treated. By "AI rate type" is meant a classification of AIs based on their typical rates of use: high rate, medium rate, and normal rates of use. An explanation

of the more specific type of use for trimmed means, called AI groups, is given in section 5.4.1 (page 18).

The median and trimmed mean criteria are based on previously reported uses. The data set for determining outliers consists of log rates of use for all PUR records during a period of time from 1 to 10 years with the same AI, unit treated, and type of record. The number of years to include in a set depends on the number of records per year in the set. The most recent records should be used because rates of use may have changed over time. However, if there are few records in the latest year, it would be better to include records from earlier years creating a larger set so that unusually high rates could be determined more reliably. The procedure used was as follows: if there were more than 10,000 records for a set in the latest year, include all records for just that year; else include records from the latest years, up to 10 years, until the total number of records was greater than 10,000; else, if there were less than 10,000 records in the last 10 years, include all records for the last 10 years.

A problem also arises with products that contain more than one AI where some of the AIs comprise a small percent of the product. Use rates of AIs with a low percent in a product will likely be less than the rates of the same AIs in products in which they are major components. To avoid this problem, the sets of uses for an AI included only records in which the percent AI in a product was 10% or greater than that of the main AI (other than inert ingredients) in the product.

5.1 Pesticide label rates

Outlier criterion 1, using pesticide label rates, would be useful, but, unfortunately, these data are not available in a DPR database for most pesticides. However, a few label rates have been entered into the database, and these are used in a limited way. In cases where a maximum label rate is available, a record would be flagged only if its rate was greater than 1.1 times the maximum label rate and at least one of the other outlier limits.

5.2 Fixed limits

Criterion 2, using fixed limits, is similar to the current fixed limit criterion, but it sets fixed limits for more situations. The current fixed limit criterion was defined only for treated units in acres and for two groups of AIs (fumigants and all others). The new criterion sets limits for all treated units (acres, cubic feet, pounds, and miscellaneous) and for both types of rates of use (pounds of AI per unit treated and pounds of AI per application). Also, rather than two groups of AIs with similar rates, four groups of AIs are defined by their typical pounds per unit treated and by whether they were adjuvants or not (Table 1) and six groups of AIs are defined by their pounds per application (Tables 3 and 4). Although AIs were classified into three groups by their pounds per acre these same groups of AIs were used for other units treated. These different groups of AIs are referred to as "AI rate types."

The three AI rate types for non-adjuvant AIs defined by pounds per acre were determined primarily by the percent of PUR agricultural records for each AI that were greater than 100 pounds per acre, which is shown for the high rate AIs and the medium rate AIs in the third column of Table 1. In most cases, a non-adjuvant AI was classified as a high rate AI if the percent of records for the AI was greater than 25%, as a medium rate AI if the percent was between 0.2 and 25, and as a normal rate AI otherwise. However, some AIs were classified as normal rate AIs even though the percent of records with rates greater than 100 was greater than 0.2, either because there were few records for an AI or because most records had low rates. Adjuvant AIs were put into their own group. The AIs in the pounds per application groups were determined similarly.

Table 1: *AIs in two of the four AI rate types for rates in pounds per acre. Listed are high rate and medium rate AIs. All other AIs are classified as either adjuvants or normal rates AIs. The number in the third column is the percent of agricultural records for that AI with rate of use greater than 100 pounds per acre from 2010 to 2014.*

Code	Active Ingredient	Pct > 100 lbs/acre
	High Rate AIs	
136	CHLOROPICRIN	93.48
385	METHYL BROMIDE	79.24
573	1,3-DICHLOROPROPENE	78.57
616	METAM-SODIUM	77.97
970	POTASSIUM N-METHYLDITHIOCARBAMATE	59.70
233	DAZOMET	47.42
	Medium Rate AIs	
270	ETHYLENE	11.47
5785	SODIUM CARBONATE PEROXYHYDRATE	3.51
2273	SODIUM TETRATHIOCARBONATE	2.98
99	CALCIUM HYDROXIDE	2.92
7	DAMINOZIDE	0.59
401	MINERAL OIL	0.47
2106	PETROLEUM DISTILLATES, REFINED	0.45
2629	KAOLIN	0.41
2210	FOSETYL-AL	0.37
1794	HYDROGEN PEROXIDE	0.36
765	PETROLEUM OIL, UNCLASSIFIED	0.30
1596	POTASH SOAP	0.25
560	SULFUR	0.22
358	LIME-SULFUR	0.20

Three fixed outlier limits were set for different kinds of PUR records determined by record type, unit treated, AI rate type, and, for non-agricultural records, by whether the application was made on a water site (Tables 2 and 5). Water sites includes the PUR sites "water area" (code 65000) and "industrial water" (code 65503). These limits were chosen somewhat subjectively by what appeared to be likely high values based on distributions of rates for each situation. More details are given on how these limits were determined in the section on criteria evaluation (section 6, page 25).

Table 2: *The fixed outlier limits of rates (pounds of AI per unit treated) for different kinds of PUR records, which are determined by record type, unit treated, AI rate type, and site type. The record types are production agricultural ("Ag") and all other types of applications, mostly non-agricultural ("Non-Ag"). The units treated are acres, cubic feet, pounds, and miscellaneous (such as bins, boxes, trees, etc.). The AI rate types are normal rate, medium rate, high rate, and adjuvant. For non-agricultural records limits are defined separately for water and non-water sites. The AIs classified as medium and high rate AIs are displayed in Table 1; all other non-adjuvant AIs are classified as normal rate AIs. Three limits are defined for each situation (fixed1, fixed2, fixed3).*

Record Type	Unit Treated	AI Rate Type	fixed1	fixed2	fixed3
Ag	Acres	Normal	50	100	150
Ag	Acres	Medium	200	500	1,000
Ag	Acres	High	1,000	2,000	5,000
Ag	Acres	Adjuvant	50	100	200
Ag	Cubic Feet	Normal	5	10	15
Ag	Cubic Feet	Medium	5	10	15
Ag	Cubic Feet	High	5	10	15
Ag	Pounds	Normal	20	50	150
Ag	Pounds	Medium	20	50	150
Ag	Pounds	High	20	50	150
Ag	Miscellaneous	Normal	20	30	50
Ag	Miscellaneous	Medium	100	200	300
Ag	Miscellaneous	High	300	500	1000
Non-Ag	Acres	Normal	2,000	5,000	100,000
Non-Ag	Acres	Medium	2,000	5,000	100,000
Non-Ag	Acres	High	2,000	5,000	100,000
Non-Ag	Acres	Water Site	5,000	20,000	200,000
Non-Ag	Cubic Feet	Normal	20	50	500
Non-Ag	Cubic Feet	Medium	20	50	500
Non-Ag	Cubic Feet	High	20	50	500
Non-Ag	Cubic Feet	Water Site	5,000	10,000	50,000
Non-Ag	Pounds	Normal	5	20	100
Non-Ag	Pounds	Medium	5	20	100
Non-Ag	Pounds	High	5	20	100
Non-Ag	Pounds	Water Site	5	20	100
Non-Ag	Miscellaneous	Normal	5,000	20,000	50,000
Non-Ag	Miscellaneous	Medium	5,000	20,000	50,000
Non-Ag	Miscellaneous	High	5,000	20,000	50,000
Non-Ag	Miscellaneous	Water Site	70,000	200,000	500,000

Table 3: *AIs in three of the six AI rate types determined by their typical rates in pounds per application for non-agricultural records. "High Lbs AI 1" include the AIs with the highest rates, "High Lbs AI 2" are the AIs with the next highest rates, and "High Lbs AI 3" are the AIs with the third highest rates. The AIs in "High Lbs AI 4" are given in Table 4, and all other non-adjuvant AIs are classified as "High Lbs AI 5."*

CODE	ACTIVE INGREDIENT
	HIGH LBS AI 1
131	CHLORINE
1881	OCTHILINONE
	HIGH LBS AI 2
161	COPPER SULFATE (PENTAHYDRATE)
417	NABAM
547	SODIUM METABISULFITE
548	SODIUM DIMETHYL DITHIO CARBAMATE
631	ARSENIC PENTOXIDE
1007	CALCIUM CARBONATE
1103	SODIUM BROMIDE
1188	CHROMIC ACID
1762	COPPER AMMONIUM CARBONATE
2148	SODIUM CHLORITE
2207	CARBON DIOXIDE
2231	COPPER OXIDE (IC)
2238	HYDROGEN CYANAMIDE
2267	NITROGEN, LIQUIFIED
4035	TETRAKIS (HYDROXYMETHYL) PHOSPHONIUM SULFATE
5785	SODIUM CARBONATE PEROXYHYDRATE
5833	5,5-DIMETHYLHYDANTOIN
	HIGH LBS AI 3
3	ACROLEIN
8	ETHYL ALCOHOL
40	ARSENIC ACID
109	CARBON TETRACHLORIDE
135	CHLORONEB
139	GLUTARALDEHYDE
231	DIURON
233	DAZOMET
248	ORTHO-PHENYLPHENOL, SODIUM SALT
250	1-(3-CHLOROALLYL)-3,5,7-TRIAZA-1- AZONIA ADAMANTANE CHLORIDE
264	EPTC
295	FORMALDEHYDE
326	CALCIUM HYPOCHLORITE
385	METHYL BROMIDE
407	TRICHLORO-S-TRIAZINETRIONE
502	PROMETRYN
508	PROPYLENE OXIDE
536	SODIUM CHLORATE
537	SODIUM FLUORIDE
539	SODIUM HYPOCHLORITE
565	BUTYLATE
573	1,3-DICHLOROPROPENE
616	METAM-SODIUM
618	SULFURYL FLUORIDE
629	ZIRAM
666	ZINC OXIDE
689	SODIUM METABORATE
970	POTASSIUM N-METHYLDITHIOCARBAMATE
1198	FREE FATTY ACIDS AND/OR AMINE SALTS
1356	ENDOTHALL, DIPOTASSIUM SALT
1615	COPPER TRIETHANOLAMINE COMPLEX
1749	2,2-DIBROMO-3-NITRILOPROPIONAMIDE
1847	ALKYL (60%C14, 30%C16, 5%C12, 5%C18) DIMETHYLBENZYL AMMONIUM CHLORIDE
1853	ALKYL (67%C12, 25%C14, 7%C16, 1%C8,C10,C18) DIMETHYLBENZYL AMMONIUM CHLORIDE
1972	ALKYLAMINO-3-AMINOPROPANE HYDROXYACETATE, ALKYL DERIVED FROM COCONUT OIL FATTY ACIDS
2056	ENDOTHALL, MONO [N,N-DIMETHYL ALKYLAMINE] SALT
2273	SODIUM TETRATHIOCARBONATE
2481	CORN GLUTEN MEAL
2629	KAOLIN
4013	SODIUM METABORATE TETRAHYDRATE
5068	ALKYL(42%C12, 26%C18, 15%C14, 8%C16, 5%C10, 4%C8)1,3-PROPYLENEDIAMINE

Table 4: *AIs in fourth AI rate type for rates in pounds per application, "High Lbs AI 4". Classification of AIs is determined by their typical pounds of AI per application in non-agricultural records.*

Code	Active Ingredient
1	TEMEPHOS
18	AMETRYNE
21	AMMONIUM SULFAMATE
25	AMMONIUM THIOCYANATE
34	MSMA
37	1,3-DICHLORO-5,5-DIMETHYLHYDANTOIN
45	ATRAZINE
53	BENEFIN
63	FENTHION
66	2-(2-(P-(DIISOBUTYL) PHENOXY) ETHOXY) ETHYL DIMETHYLBENZYL AMMONIUM DIMETHYLBENZYL AMMONIUM C
70	BENSULIDE
79	BORAX
83	BROMACIL
88	TRICHLORFON
89	2-BUTOXYETHANOL
99	CALCIUM HYDROXIDE
104	CAPTAN
105	CARBARYL
106	CARBOFURAN
108	CARBON DISULFIDE
112	DICHLOBENIL
130	CHLORDANE
136	CHLOROPICRIN
141	CHLORPROPHAM
142	CITRIC ACID
151	COPPER HYDROXIDE
153	COPPER NAPHTHENATE
158	COPPER OXYCHLORIDE SULFATE
162	COPPER SULFATE (BASIC)
171	CREOSOTE
173	CRYOLITE
175	COPPER OXIDE (OUS)
179	CHLORTHAL-DIMETHYL
185	1,2-DICHLOROPROPANE, 1,3-DICHLOROPROPENE AND RELATED C3 COMPOUNDS
187	DDVP
189	DEET
190	S,S,S-TRIBUTYL PHOSPHOROTRITHIOATE
195	DIATOMACEOUS EARTH
198	DIAZINON
205	SODIUM DICHLORO-S-TRIAZINETRIONE
211	MANCOZEB
216	DIMETHOATE
228	DIPHENYLAMINE
229	DIQUAT DIBROMIDE
253	CHLORPYRIFOS
256	ANILAZINE
259	ENDOSULFAN
262	ENDRIN
263	EPN
268	ETHION
271	ETHYLENE DIBROMIDE
276	ETHYLENE GLYCOL
305	TETRACHLORVINPHOS
317	HEPTACHLOR
329	HYDROGEN CHLORIDE
335	PHOSMET
358	LIME-SULFUR
359	LINDANE
365	MAGNESIUM CHLORIDE
367	MALATHION
369	MANEB
374	MCPP
389	FATTY ACIDS, METHYL ESTERS
392	METHYL ISOTHIOCYANATE
394	METHYL PARATHION
401	MINERAL OIL
404	ETHOPROP
418	NALED
445	PROPARGITE
448	ORTHO-PHENYLPHENOL
455	PARA-DICHLOROBENZENE
459	PARATHION
464	PCNB
465	PENTACHLOROPHENOL
473	PETROLEUM HYDROCARBONS
478	PHORATE
480	MEVINPHOS
484	ALUMINUM PHOSPHIDE
488	PIPERALIN
503	PROPANIL
504	PROPAZINE
505	PROPIONIC ACID
509	PYRAZON
516	CYCLOATE
522	ORTHO-BENZYL-PARA-CHLOROPHENOL
531	SIMAZINE
560	SULFUR
561	SULFUR DIOXIDE
569	TRIBUTYLTIN OXIDE
575	ALDICARB
587	THIABENDAZOLE
589	THIRAM
597	TRIFLURALIN
603	SIDURON
624	ZINC CHLORIDE
636	2,4-D
662	UREA
677	CHLOROTHALONIL
678	ALACHLOR
688	SODIUM CYANIDE
694	PROPYZAMIDE
696	SODIUM NITRATE
714	COPPER
721	SODIUM CHLORIDE
748	ALPHA-ALKYLARYL-OMEGA-HYDROXYPOLY(OXYETHYLENE)
763	PETROLEUM DISTILLATES
765	PETROLEUM OIL, UNCLASSIFIED
769	BORIC ACID
774	SODIUM DODECYLBENZENE SULFONATE
778	MALEIC HYDRAZIDE, DIETHANOLAMINE SALT
786	MCPA, DIMETHYLAMINE SALT
787	MCPA, ISOOCTYL ESTER
801	2,4-D, ALKANOLAMINE SALTS (ETHANOL AND ISOPROPANOL AMINES)
802	2,4-D, BUTOXYETHANOL ESTER
805	2,4-D, DIETHANOLAMINE SALT
806	2,4-D, DIMETHYLAMINE SALT
809	2,4-D, ISOOCTYL ESTER
810	2,4-D, ISOPROPYL ESTER
816	2,4-D, TRIISOPROPYLAMINE SALT
834	BROMOXYNIL OCTANOATE
838	4-(2,4-DB), DIMETHYLAMINE SALT
862	XYLENE RANGE AROMATIC SOLVENT
892	AMMONIUM THIOSULFATE
905	SODIUM BISULFATE
933	ACETIC ACID
935	ORTHO-PHENYLPHENOL, POTASSIUM SALT
936	ORTHO-BENZYL-PARA-CHLOROPHENOL, POTASSIUM SALT
971	TCMTB
979	LIMONENE
991	THYMOL
1015	COTTONSEED OIL
1040	POLYBUTENES
1047	DODECYL GUANIDINE HYDROCHLORIDE
1074	MENTHOL
1105	TRIS (HYDROXYMETHYL) NITROMETHANE
1134	SODIUM BICARBONATE
1138	2,4-D, TRIETHYLAMINE SALT
1159	MODIFIED PHTHALIC GLYCEROL ALKYD RESIN
1171	HEXAHYDRO-1,3,5-TRIS (2-HYDROXYETHYL)-S-TRIAZINE
1173	ALPHA-ALKYLPHENYL-OMEGA-HYDROXYPOLY(OXYETHYLENE)
1178	POLYOXYETHYLENE SORBITOL, MIXED ETHER ESTER
1212	ABIETIC ANHYDRIDE
1213	DIETHYLENE GLYCOL ABIETATE
1214	PARAFFIN WAX
1221	SESAME OIL
1231	VEGETABLE OIL
1244	ALPHA-OCTYLPHENYL-OMEGA-HYDROXYPOLY(OXYETHYLENE)
1254	HEXYLENE GLYCOL
1314	POLY-I-PARA-MENTHENE
1363	AMMONIUM SULFATE
1382	2,4,5-TRICHLOROPHENOL
1399	BENTONITE
1512	CHLORMEQUAT CHLORIDE
1573	BROMACIL, LITHIUM SALT
1576	POLY(OXYETHYLENE) (DIMETHYLIMINO) ETHYLENE (DIMETHYLIMINO) ETHYLENE (DIMETHYLIMINO) ETHYLENE D
1596	POTASH SOAP
1600	POLYETHYLENE GLYCOL
1601	PARAQUAT DICHLORIDE
1605	ALPHA-(PARA-TERT-BUTYLPHENYL)-OMEGA-HYDROXYPOLY (OXYETHYLENE) PHOSPHATE
1626	ETHEPHON
1640	CYANAZINE
1641	ISOPARAFFINIC HYDROCARBONS
1656	2,4,5-TRICHLOROPHENOL, SODIUM SALT
1664	1-(2-HYDROXYETHYL)-2-ALKYL-2-IMIDAZOLINE, ALKYL
1673	SODIUM CACODYLATE
1682	DIDECYL DIMETHYL AMMONIUM CHLORIDE
1684	THIOPHANATE
1689	METHIDATHION
1692	METRIBUZIN
1696	THIOPHANATE-METHYL
1697	METHAMIDOPHOS
1698	POLYOXYETHYLENE POLYOXYPROPYLENE
1709	OCTYL DECYL DIMETHYL AMMONIUM CHLORIDE
1710	DIOCTYL DIMETHYL AMMONIUM CHLORIDE
1717	NAPHTHA, HEAVY AROMATIC
1728	NAPROPAMIDE
1743	ALPHA-(PARA-NONYLPHENYL)-OMEGA-HYDROXYPOLY (OXYETHYLENE)
1755	CARBOXIN
1766	ALPHA-ALKYL (C12-C14)-OMEGA-HYDROXYPOLY(OXYETHYLENE)
1768	LIGNIN SULFONIC ACID, ZINC SALT
1770	LIGNIN SULFONIC ACID, MANGANESE SALT
1778	COPPER SULFATE (ANHYDROUS)
1794	HYDROGEN PEROXIDE
1800	DISODIUM OCTABORATE TETRAHYDRATE
1808	SODIUM TETRABORATE (PENTAHYDRATE)
1810	TEBUTHIURON
1814	PETROLEUM DISTILLATES, AROMATIC
1818	SODIUM DICHLORO-S-TRIAZINETRIONE DIHYDRATE
1844	ALPHA-ALKYL (C10-C14)-OMEGA-HYDROXYPOLY(OXYETHYLENE)
1846	ALKYL (50%C14, 40%C12, 10%C16) DIMETHYLBENZYL AMMONIUM CHLORIDE
1855	GLYPHOSATE, ISOPROPYLAMINE SALT
1857	FENAMIPHOS
1859	POTASSIUM PEROXYMONOSULFATE
1868	ORYZALIN
1871	HEXAZINONE
1884	ALPHA-ALKYL (C9-C16)-OMEGA-HYDROXYPOLY(OXYETHYLENE)
1900	ETHOFUMESATE
1910	OXAMYL
1929	PENDIMETHALIN
1930	DIFENZOQUAT METHYL SULFATE
1933	THIOBENCARB
1934	POTASSIUM DIMETHYL DITHIO CARBAMATE
1952	THIABENDAZOLE, HYPOPHOSPHITE SALT
1973	OXYFLUORFEN
1996	METOLACHLOR
2008	PERMETHRIN
2016	AMITRAZ
2017	OXADIAZON
2019	NORFLURAZON
2029	DAZOMET, SODIUM SALT
2034	DICLOFOP-METHYL
2038	5-CHLORO-2-METHYL-4-ISOTHIAZOLIN-3-ONE
2039	2-METHYL-4-ISOTHIAZOLIN-3-ONE
2043	POLYMERIZED PINENE
2045	PETROLEUM OIL, PARAFFIN BASED
2069	FATTY ACIDS, MIXED
2071	KEROSENE
2080	1-BROMO-3-CHLORO-5,5-DIMETHYL HYDANTOIN
2081	IPRODIONE
2084	IMAZALIL
2085	MAGNESIUM PHOSPHIDE
2088	COCONUT OIL SOAP
2106	PETROLEUM DISTILLATES, REFINED
2108	CUPROUS THIOCYANATE
2123	ALPHA-ALKYL (C12-C18)-OMEGA-HYDROXYPOLY(OXYETHYLENE) POLY(OXYPROPYLENE)
2129	VINCLOZOLIN
2131	TRICLOPYR, TRIETHYLAMINE SALT
2132	METALAXYL
2138	CYANURIC ACID
2149	SULFOMETURON-METHYL
2154	ACID BLUE 9, DIAMMONIUM SALT
2166	ETHALFLURALIN
2167	POTASSIUM DICHROMATE
2170	TRICLOPYR, BUTOXYETHYL ESTER
2171	CYPERMETHRIN
2177	SETHOXYDIM
2179	TRIBUTYLTIN METHACRYLATE

Table 4: (continued) *AIs classified as "High Lbs AI 4" by pounds of AI per application*

Code	Active Ingredient
2210	FOSETYL-AL
2216	ALPHA-ISOOCTADECYL-OMEGA-HYDROXYPOLY(OXYETHYLENE)
2236	PRODIAMINE
2249	CLOFENTEZINE
2251	Z-8-DODECENYL ACETATE
2257	IMAZAPYR, ISOPROPYLAMINE SALT
2259	PACLOBUTRAZOL
2263	BENSULFURON METHYL
2264	1,3-DICHLORO-5-ETHYL-5-METHYLHYDANTOIN
2266	OLEIC ACID, METHYL ESTER
2270	UREA DIHYDROGEN SULFATE
2279	FLURIDONE
2286	CYROMAZINE
2289	ISOXABEN
2291	PEROXYACETIC ACID
2295	BRONOPOL
2301	GLYPHOSATE, MONOAMMONIUM SALT
2305	FLUTOLANIL
2307	TRIADIMENOL
2308	DITHIOPYR
2322	CORN PRODUCT, HYDROLYZED
2326	MCPA
2327	GLYPHOSATE-TRIMESIUM
2334	DIETHYLENE GLYCOL
2335	SOYBEAN OIL
2341	IMAZETHAPYR, AMMONIUM SALT
2375	ALKYL (61%C12,23%C14, 11%C16,2.5%C8 & C10, 2.5%C18) DIMETHYL 2.5%C18) DIMETHYL B
2409	BACILLUS SPHAERICUS, SEROTYPE H-5A5B, STRAIN 2362
2601	HYDROTREATED PARAFFINIC SOLVENT
2739	NONANOIC ACID
2768	PETROLEUM DISTILLATES, ALIPHATIC
2777	PHOSPHORIC ACID, MONOPOTASSIUM SALT
2800	POLYETHYLENE GLYCOL OLEATE
2814	POLYOXYETHYLENE SORBITAN TRIOLEATE
2997	GLYPHOSATE
3208	FORMIC ACID
3264	LECITHIN
3338	POLYOXYETHYLENE SOYBEAN OIL FATTY ACID ESTER
3395	SODIUM DICHROMATE
3519	METHYLATED SOYBEAN OIL
3537	CLOMAZONE
3541	PHOSPHINE
3547	COPPER CITRATE CHELATE
3548	COPPER GLUCONATE CHELATE
3549	COPPER ETHYLENEDIAMINE COMPLEX
3551	COPPER ETHANOLAMINE COMPLEXES, MIXED
3566	CLETHODIM
3628	CORN SYRUP
3738	POLYMERIZED ACRYLIC ACID
3779	QUILLAJA
3829	NICOSULFURON
3833	OIL OF JOJOBA
3850	TEBUCONAZOLE
3855	ALKYL (95%C14, 3%C12, 2%C16) DIMETHYL BENZYL AMMONIUM CHLORIDE
3857	BACILLUS THURINGIENSIS (BERLINER), SUBSP. ISRAELENSIS, SEROTYPE H-ISRAELENSIS, SEROTYPE H-1
3905	FENBUCONAZOLE
3929	ALPHA-UNDECYL-OMEGA-HYDROXYPOLY(OXYETHYLENE)
3933	OLEIC ACID, ETHYL ESTER
3940	PYRITHIOBAC-SODIUM
3946	GLUFOSINATE-AMMONIUM
3966	MYROTHECIUM VERRUCARIA, DRIED FERMENTATION SOLIDS & SOLUBLES, STRAIN AARC-SOLUBLES, STRAIN AARC-0
3976	N,N-BIS-(2-(OMEGA-HYDROXYPOLY (OXYETHYLENE)/POLY (OXYPROPYLENE))ETHYL) (OXYPROPYLENE))ETHYL)A
3979	CLARIFIED HYDROPHOBIC EXTRACT OF NEEM OIL
3984	THIAZOPYR
3996	ALPHA-PINENE BETA-PINENE COPOLYMER
4000	CYPRODINIL
4006	ALKYL (C9-C11) OLIGOMERIC D-GLUCOPYRANOSIDE
4015	ALKYL (C8,C10) POLYGLUCOSIDE
4022	PROPAMOCARB HYDROCHLORIDE
4025	METHYLATED FATTY ACIDS FROM CANOLA OIL
5007	DIGLYCOLAMINE SALT OF 3,6-DICHLORO-O-ANISIC ACID
5015	AMMONIUM PROPIONATE
5016	ALPHA-(PARA-NONYLPHENYL)-OMEGA-HYDROXYPOLY (OXYETHYLENE), PHOSPHATE (OXYETHYLENE), PHOSPHATE E
5022	E-11-TETRADECEN-1-YL ACETATE
5023	Z-11-TETRADECEN-1-YL ACETATE
5024	DIFENOCONAZOLE
5036	BROMOXYNIL HEPTANOATE
5037	POTASSIUM BICARBONATE
5050	CLOPYRALID, MONOETHANOLAMINE SALT
5067	1,4-DIMETHYLNAPHTHALENE
5133	S-METOLACHLOR
5179	ORCHEX 796 OIL
5181	DIMETHYL DICOCOALKYL AMMONIUM SALT WITH NAPHTHALENESULFONIC ACID, NAPHTHALENESULFONIC ACID,
5204	FATTY ACIDS DERIVED FROM TALLOW
5224	EMULSIFIABLE METHYLATED VEGETABLE OIL
5230	SUCROSE OCTANOATE
5451	KRESOXIM-METHYL
5538	2,4-D, TRIISOPROPANOLAMINE SALT
5698	METHOXYFENOZIDE
5748	CYHALOFOP-BUTYL
5766	POTASSIUM PHOSPHITE
5802	FLUMIOXAZIN
5810	GLYPHOSATE, DIAMMONIUM SALT
5820	GLYPHOSATE, POTASSIUM SALT
5838	ALPHA-ALKYL (C10-C20)-OMEGA-HYDROXYPOLY(OXYETHYLENE)
5839	DIDECYL DIMETHYL AMMONIUM CARBONATE
5840	DIDECYL DIMETHYL AMMONIUM BICARBONATE
5841	BACILLUS THURINGIENSIS, SUBSP. ISRAELENSIS, STRAIN AM 65-52
5843	SODIUM ALPHA-OLEFIN (C14-C16) SULFONATE
5844	ALCOHOLS, C12-C14, ALIPHATIC
5862	ALPHA-ALKYL (C9-C18)-OMEGA-HYDROXYPOLY(OXYETHYLENE)
5868	POLY(OXYETHYLENE) POLY (OXYPROPYLENE) GLYCOL MONOALLYL ETHER
5873	POLYOXYETHYLENE SORBITAN MIXED FATTY ACID ESTERS
5874	N-BUTYL-1,2-BENZISOTHIAZOLIN-3-ONE
5881	FATTY ACIDS, C16-C18 AND C18-UNSATURATED, METHYL ESTERS
5895	ALPHA-(ORTHO,PARA-DINONYLPHENYL)-OMEGA-HYDROXYPOLYOXY(ETHYLENE) HYDROXYPOLYOXY(ETHYLENE) P
5918	FLUROXYPYR, 1-METHYLHEPTYL ESTER
5928	AMINOPYRALID, TRIISOPROPANOLAMINE SALT
5963	AMMONIUM NONANOATE
5964	CHLORANTRANILIPROLE
5972	GLYPHOSATE, DIMETHYLAMINE SALT
5978	IPCONAZOLE
5983	METCONAZOLE
5998	AMINOCYCLOPYRACHLOR
6004	FLUOPYRAM
6031	SODIUM BROMOSULFAMATE, SODIUM CHLOROSULFAMATE, POTASSIUM BROMOSULFAMATE, POTASSIUM BROMOSULFAMATE,
6065	MARGOSA OIL
6102	QUINCLORAC, DIMETHYLAMINE SALT
6103	COPPER DIAMMONIUM DIACETATE COMPLEX
90189	DEET, OTHER RELATED
90480	MEVINPHOS, OTHER RELATED
90849	DICAMBA, DIMETHYLAMINE SALT, OTHER RELATED

Table 5: *The fixed outlier limits of pounds of AI per application for different kinds of non-agricultural PUR records, which are determined by AI rate type and site type. The AI rate types are adjuvant and five groups of AIs from those used at the highest amounts ("High Lbs AI1") to the lowest amounts ("High Lbs AI5"). The site types are water body and other site. Three limits are defined for each situation (fixed1, fixed2, fixed3). The AIs in each of the first three AI rate types are listed in Table 3, the AIs in the fourth type, "High Lbs AI 4", are listed in Table 4; AI rate type "High Lbs AI 5" consists of all other non-adjuvant AIs.*

AI Rate Type	Site Type	fixed1	fixed2	fixed3
High Lbs AI 5	Other Site	30	100	200
High Lbs AI 4	Other Site	1,000	3,000	5,000
High Lbs AI 3	Other Site	5,000	10,000	20,000
High Lbs AI 2	Other Site	40,000	50,000	60,000
High Lbs AI 1	Other Site	100,000	200,000	400,000
Adjuvant	Other Site	2,000	5,000	10,000
High Lbs AI 5	Water Site	1,000	2,000	3,000
High Lbs AI 4	Water Site	5,000	10,000	50,000
High Lbs AI 3	Water Site	20,000	40,000	60,000
High Lbs AI 2	Water Site	50,000	60,000	70,000
High Lbs AI 1	Water Site	150,000	250,000	500,000
Adjuvant	Water Site	10,000	20,000	40,000

5.3 Median rates times a factor

Criterion 3, median*x, is also similar to the current median*x criterion except that the new criterion is based on all rates for each AI, unit treated, and record type in a 1 to 10 year period rather than rates for each product, site, unit treated, and record type for one year. This new criterion also differs from the current criterion by setting limits for both kinds of rates (pounds per unit treated and pounds per application) and by using more multiplication factors (the "x" in the criterion name) than just 50, which was used in the current criterion.

5.4 Trimmed mean + standard deviation times a factor

Criterion 4, mean+y*SD, sets an outlier limit by adding a trimmed mean with a multiple of the standard deviation. The means and standard deviations are calculated from a set of rates in which the highest and lowest 5% values have been removed.

5.4.1 Defining AI groups

In sets of rates for some AIs, units treated, and record types, there are different kinds of uses with distinctly different rates. For example, the frequency distribution of spinosad pounds per acre, shown earlier (Figure 4), has peaks at two different rates. Using such a distribution to determine outliers with the trimmed mean criterion will likely miss many erroneous rates since it depends on the standard deviation. In this case with widely different rates, the standard deviation is quite large and results in high outlier limits. However, if this set of rates were separated into two different distributions, each distribution would have a smaller standard deviation than in the combined distribtion, and the outlier limits for each group would be smaller and more accurate.

Many factors could be used to distinguish such distinct uses, such as the pest being treated, the site or crop treated, the product used, the geographical location, or the application date. The PUR does not have the pest treated but it does have data on the other factors. However, after examining many different situations, the most important factors were found to be site treated and product used, so those are the two factors that will be used here.

Determining distinct groups of rates for the over 2,000 AIs in the PUR would be time consuming to do by eye, so a more automated procedure is needed. The procedure needs to determine if there are two or more groups of distinct uses for each AI and what sites, products, or combinations of the two, determine each group. Two procedures that could be used are regression trees (Breiman, 1984) and the Patient Rule-Induction Method (PRIM)(Friedman and Fisher, 1999). However, in trials using these procedures, neither were completely adequate for the needs here, so another method was developed combining aspects of both procedures.

The basic idea is to determine if there are two or more modes in a distribution of rates for each AI and, if so, determine whether the rates can be separated by the sites treated or products used. Because the rates of use are usually similar for similar sites, sites are grouped into general site categories, such as small fruits, citrus, pome fruits, vegetables, etc. Also, pesticide products are grouped by combining all products with the same two-part registration number (which is called "regno_short" here). Using such general sites and products provides more records that can be used to more accurately determine distinct uses. Groups are considered similar to one another if their mean rates are close to one another relative to the variances of their distributions. For example, two groups with means close to one another but with very narrow distributions so that there are few rates common to both groups, would be kept as two separate groups. On the other hand, if two groups had means far apart but with large variances, then they would be put into the same group.

The algorithm used, simplified somewhat, is as follows:

```
FOR each AI, record type, and unit treated
    Get statistics for the set of rates and store them in a new table named
    "AI stats table" with columns:
    'chem_code', 'ai_group', 'site_general', 'regno_short',
    'ago_ind', 'unit_treated', 'chemname',
    'num_recs', 'median', 'mean', 'sd_rate_orig', 'sd_rate_adjusted'.

    FOR each general site and product ("regno_short") for the current AI
        Get set of log(rates) for this site and product
        Remove the highest and lowest 5% from this set of log rates
        Calculate the following from the trimmed set of log rates:
            Number of records (num_recs)
            Median rate (median)
            Mean rate (mean)
            Standard deviation (sd_rate_orig)

        Adjust the standard deviation by increasing its value for AIs with few
        records:
            Set minimum value for sd_rate_orig (sd_min) = 0.05;
            this number was used because using a standard deviation below that
            produced a high percent of outliers.  There is no correlation of
            sd_rate_orig with mean and only a weak correlation with number of records.
            sd_adjust_a = 60
            sd_adjust_b = 1.7
            sd_rate_adjusted =
                MAX(sd_rate_orig, sd_min)*(1 + sd_adjust_a*(num_recs^(-sd_adjust_b)))

        Add this set of statistics as another row to the AI stats table.

    The following procedure will combine rates from different sites and products for
    each AI into groups with similar values of trimmed mean and separated from
    other groups. A description of how the algorithm determines whether or not sets
    of records are put into the same group is given just after this algorithm.
    Each group is identified by a number, ai_group.

    Sort the AI stats table by trimmed mean starting with the highest mean.
    REPEAT until all records in the AI stats table have been assigned an ai_group
        For the first time through set ai_group_num = 1,
            then next time through increase ai_group_num by 1.
        MOVE through each record in the AI stats table without an ai_group number
            For the first record in the AI stats table without an ai_group number,
                set its ai_group = ai_group_num
            For other records without an ai_group, but which can be grouped with those
                in the first record, set their ai_group = ai_group_num
            If there are few records for any AI stats table row,
                assign its ai_group = 0.
                Few records means less than the maximum of 10 and the number of
                records times 0.001. Later the procedure will assign these
                small groups to one of the main groups.

    Since rates from each site and product were assigned to groups sequentially,
    before all groups were determined, it may be that some groups are actually closer
    to a group created later. Also small groups have not yet been assigned to any
    group. Reassign a set of rates to another group if its trimmed mean is closer to the
    trimmed mean of that group.

    Find the trimmed mean rate for each of the existing groups, call these the
    group means
    FOR each record in the AI stats table
        Find the group mean that is closest to the mean of the current record
            and set this record's ai_group equal to the group number with that
            group mean.
```

In this algorithm the current set of rates and the next set of rates (whose statistics are given in the current and next records in the AI stats table) are put into the same group if either of the following two conditions are satisfied:

Condition 1:

$$|\overline{r_c} - \overline{r_n}| < r_d$$

where

$\overline{r_c}$ is the trimmed mean of the log rates in the current group,
$\overline{r_n}$ is the trimmed mean of the log rates in the next group, and
r_d is a parameter, currently set to 1.

Condition 2:

$$\frac{(SS_T - SS_c - SS_n)}{SS_T} < SS_d$$

where

SS_c is the sum of squares of the trimmed set of log rates in the current group,
SS_n is the sum of squares of the trimmed set of log rates in the next group,
SS_T is the sum of squares of the trimmed set of log rates in both sets, and
SS_d is a parameter, currently set to 0.6.

The sum of squares of a set of values, such as the trimmed log rates, is given by:

$$SS_i = \sum (r_i - \overline{r_i})^2$$

Condition 1 means that two sets of rates are put into one group if their means differ by less than one order of magnitude. This is similar to the criterion used in PRIM.

Condition 2 means that two sets of rates are put into one group if their distributions significantly overlap. Thus, if the means of two sets are far apart but their variances are also so large that their distributions overlap, then the two sets will be considered similar. This condition is somewhat similar to the criterion used in regression trees.

5.4.2 Results from the procedure to find AI groups

Eight examples of the results from running this procedure are given in Figures. 12 to 19. Rates of use for s-methoprene (Figure 12) fall into two distinct groups. One group, with the larger rates, consists mostly of applications on vegetables, and the second group consists mostly of applications on nuts.

Rates for methyl bromide (Figure 13) also fall into two groups. The group with larger rates consists mostly of applications on small fruits and nurseries, and the smaller rate group is mostly applications on nuts and stone fruits. Although some of the rates used on nuts are similar to those used on small fruits and nurseries, most rates are much lower and so are included in a separate group.

Pyriproxyfen has two groups (Figure 14), but these groups are determined primarily by the pesticide products used. One group consists of applications with ant baits, and the second group consists of applications by all other products. Baits are commonly used at lower rates than other kinds of formulations.

Spinosad is separated into three groups (Figure 15), also determined primarily by the products applied. In this case the two lower rate groups are from different baits and the higher rate group is from all other products.

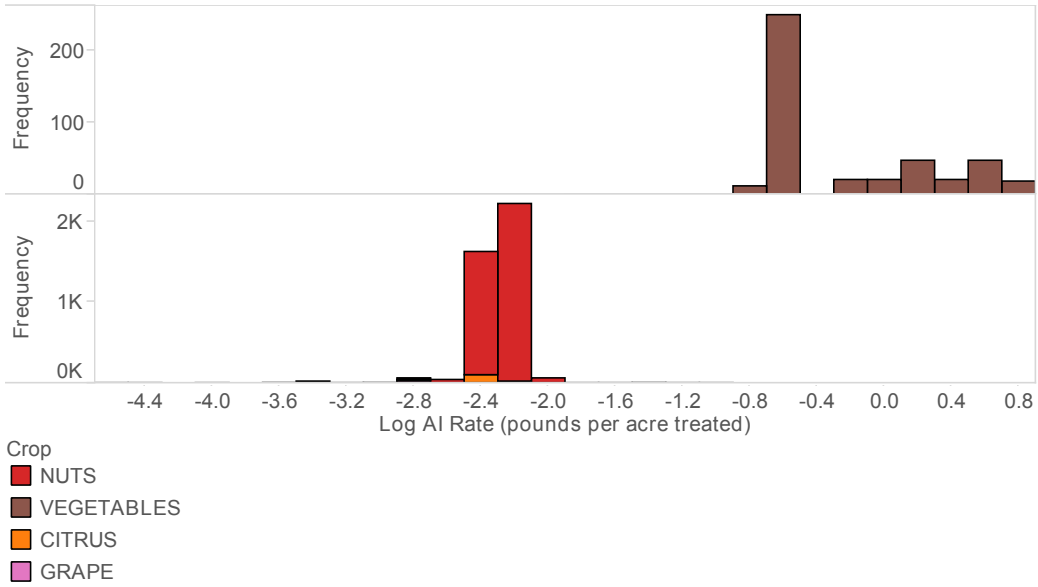

Figure 12: *Frequency distribution of log rates of use on different groups of s-methoprene applications.*

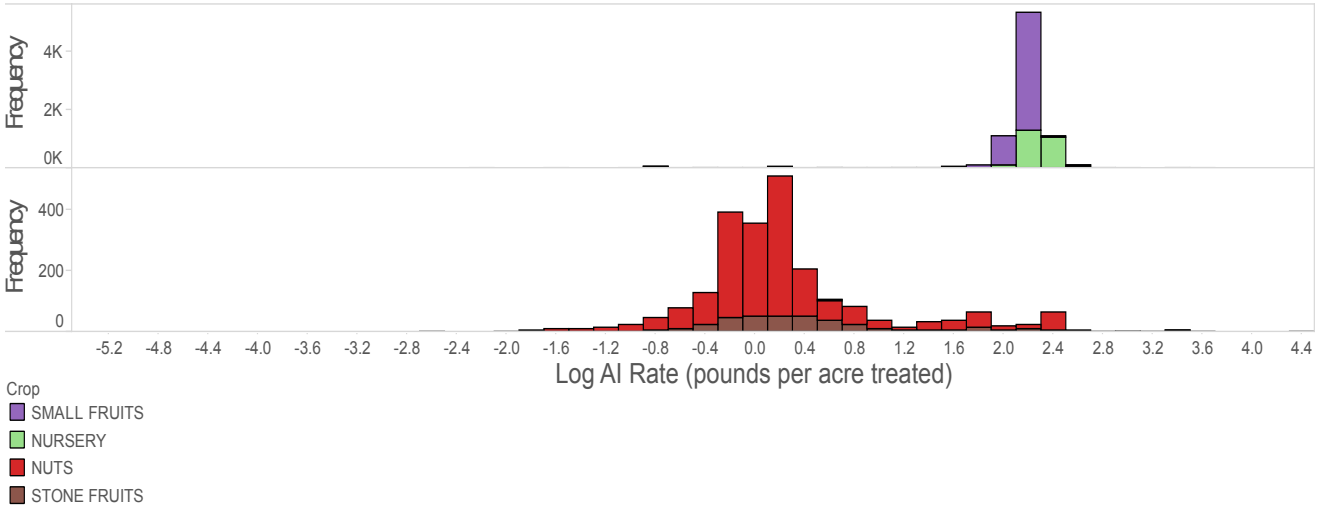

Figure 13: *Frequency distribution of log rates of use on different groups of methyl bromide applications.*

This procedure also works well for rates in pounds per application. For example, abamectin uses are clearly divided into two main groups of uses: landscape maintenance and structural pest control (Figure 16). Borax is used at different rates in structural pest control, rights of way, and forests (Figure 17). As a final example, ethephon uses are distinguished by two different kinds of products: some products, such as Proxy, are used to suppress seedheads and growth of some grasses; and other products, such as Florel Brand Growth Regulator and Florel Brand Fruit Eliminator, are used to prevent fruit growth on ornamental trees (Figure 18).

Finally, in some cases, none of the information in the PUR is useful in determining what distinguishes

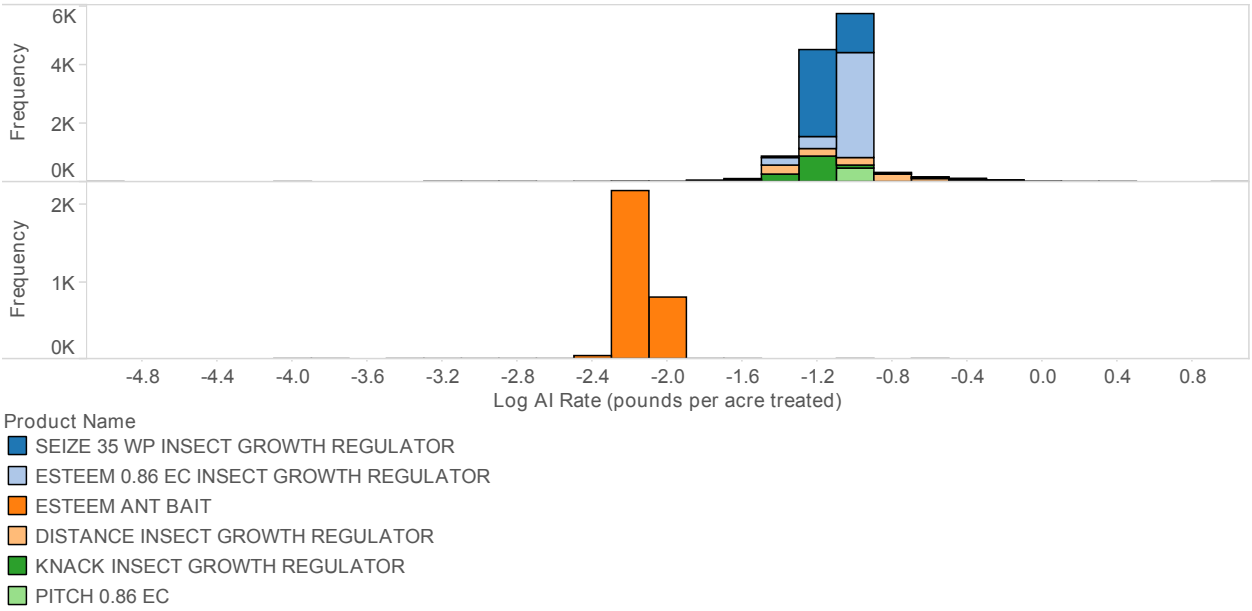

Figure 14: *Frequency distribution of log rates of use on different groups of pyriproxyfen applications.*

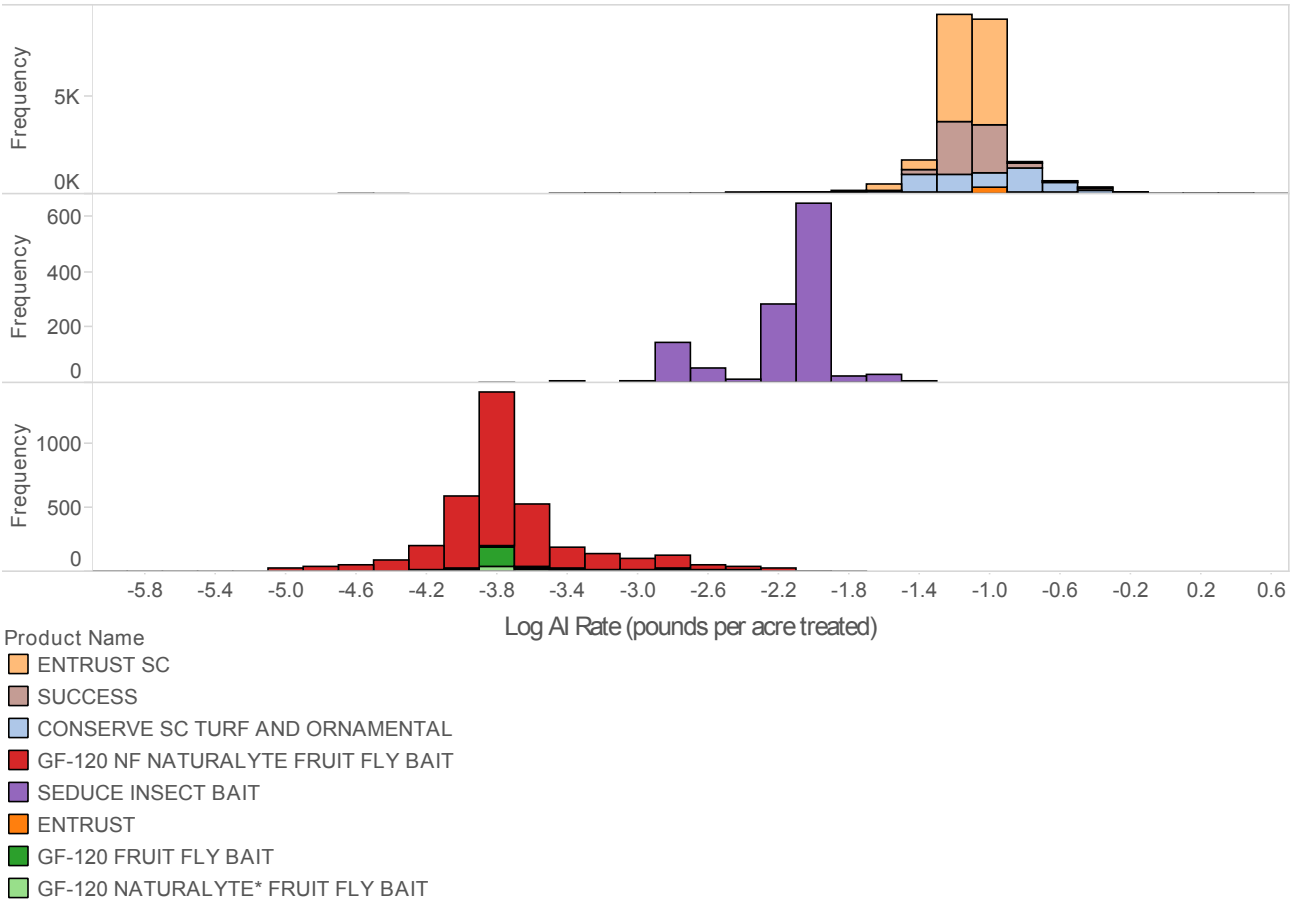

Figure 15: *Frequency distribution of log rates of use on different groups of spinosad applications.*

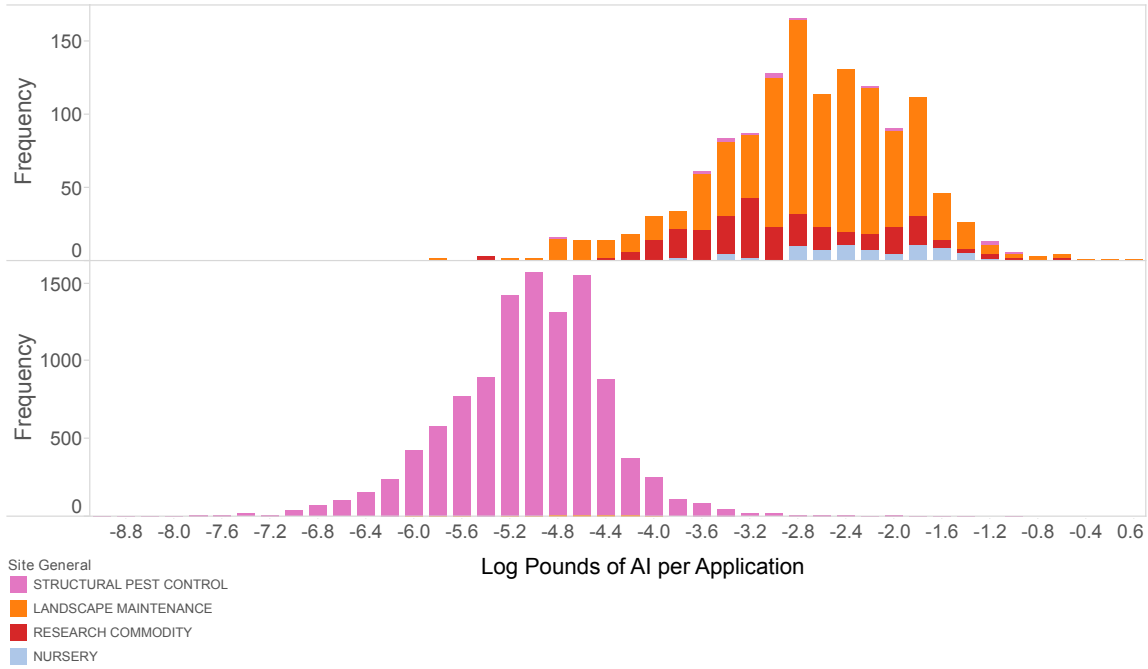

Figure 16: *Frequency distribution of log pounds AI per application on different groups of non-agricultural abamectin applications.*

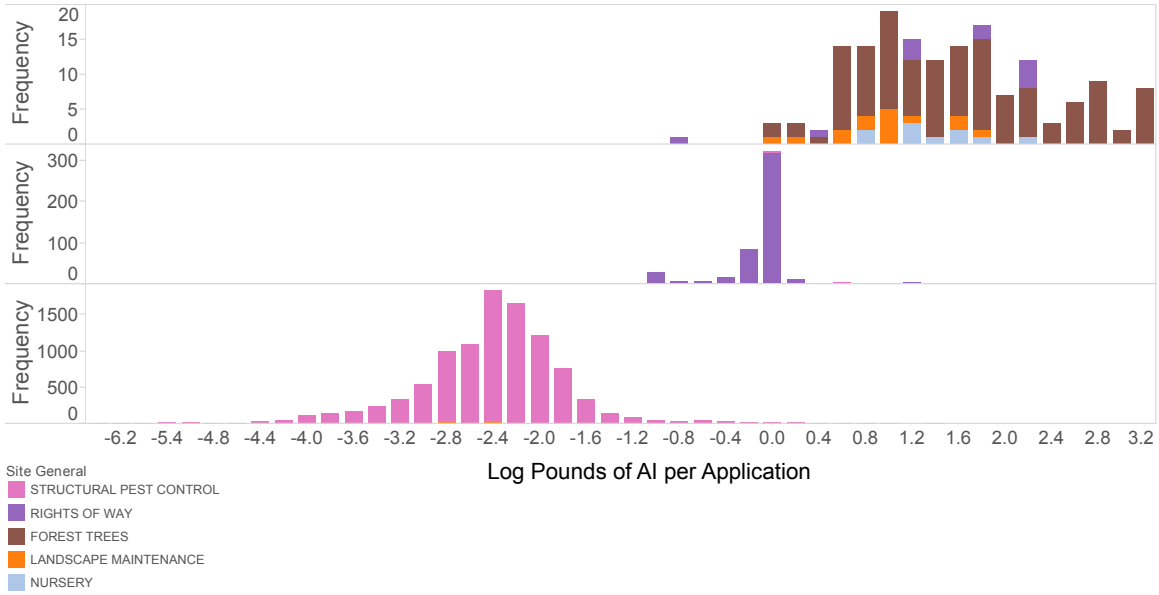

Figure 17: *Frequency distribution of log pounds AI per application on different groups of non-agricultural borax applications.*

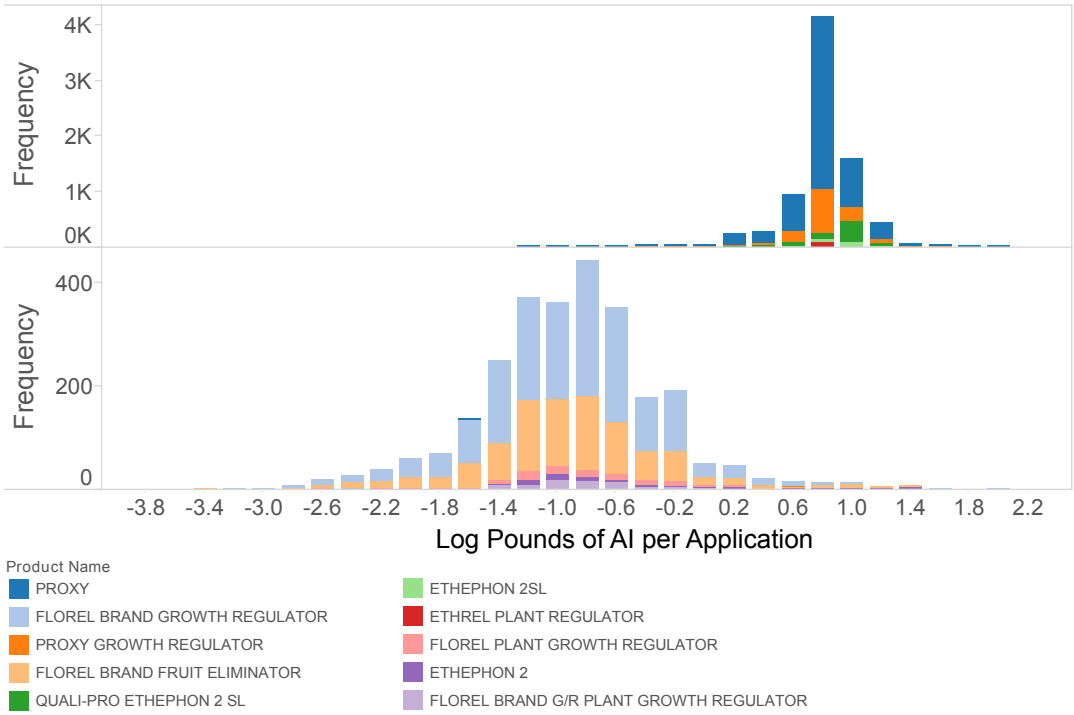

Figure 18: *Frequency distribution of log pounds AI per application on different groups of non-agricultural ethephon applications.*

different groups of AIs. For example, there are two groups in dithiopyr rates (Figure 19), but none of the PUR variables were able to separate these two groups. Nearly all applications are made to nurseries, so the site treated does not help in distinguishing the two groups. Even though the lower group is primarily due to one product, there are many other applications of this product at higher rates and there is nothing about that product that distinguishes it from other products. However, this situation is rare.

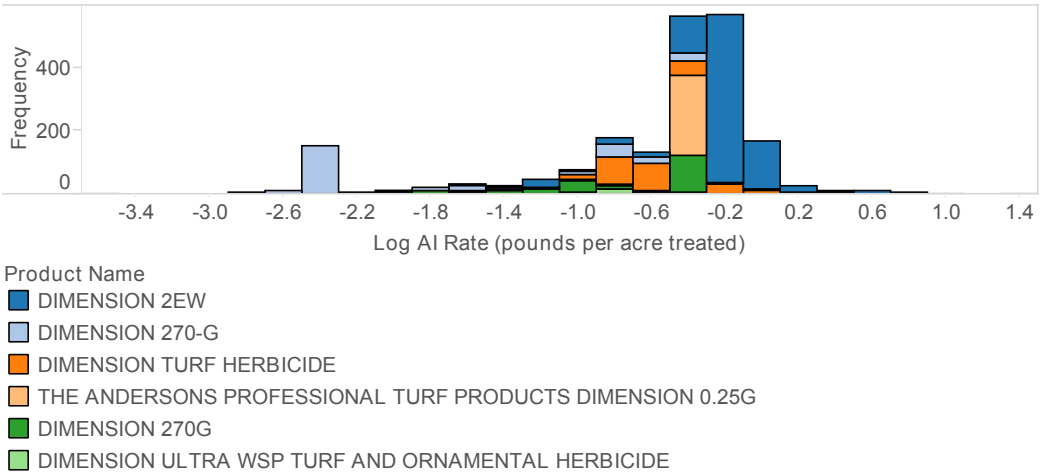

Figure 19: *Frequency distribution of log rates of use for dithiopyr applications.*

6 Criteria evaluation

There are many methods for identifying outliers, and for each of these methods there are many choices for parameter values, such as the parameter that is multiplied by the standard deviation in criterion 4. This section discusses four different methods to evaluate these criteria. Since it is not known which rates are correct or not, all these methods must depend partly on subjective judgments. But if some criterion appears best using different evaluation methods, one can have reasonable confidence that it is reliable.

One method for evaluating the criteria is to examine many graphs of distributions of rates with the outlier limits marked and subjectively judge how well they do. Another is to determine how many records are flagged by each criterion. A third method compares the rankings of the limits to one another. A fourth method compares the outlier limits with known maximum label rates.

6.1 Visual inspections of distributions with the different outlier limits

First, several graphs of frequency distributions of log rates of use will be presented for different AIs or AI groups (Figures 20 – 23). There are four pages of graphs, and each page displays four graphs with a similar rate measure, record type, and unit treated. The top graphs on each page have the narrowest distributions, and each graph below has progressively broader distributions. These distributions were chosen to be representative of all the different kinds of distributions. Vertical lines are drawn on each graph at different outlier limit values. Also, lines at the median, 99th percentile, and 99.9th percentile values are drawn as reference. The different vertical lines on these graphs are described in Table 6.

For very narrow distributions of rates of use, the median limits are probably too high, and the fixed limits are also usually too high (Figures 20 – 22). This is not surprising since the median criteria does not account for the variance in rates. The best limits in most cases are probably mean+10*SD or mean+12*SD. This is a common kind of distribution and all look very similar to these example graphs.

As the distributions become broader, the trimmed mean limits naturally become higher, usually fairly close to the median limits and sometimes the fixed limits (Figures 20 – 23). These distributions are more heterogeneous and the best limits are often different for different situations. However, in most cases the best limits are the trimmed means, mean+5*SD to mean+10*SD. Most distributions of agricultural records with units in acres are narrower than the distributions for other kinds of records. Different median limits are used for agricultural records with units in acres than all other kinds of records because these other kinds of records include widely different kinds of applications. For example, miscellaneous units can refer to many different things, such as individual plants, tree holes, bins, pallets, etc., and for non-agricultural records one application could mean anything from a spot treatment on a lawn to an entire building. For agricultural records in acres, the median limits displayed are median*50, median*100, and median*200 and for the other kinds of records the median limits are median*200, median*5000, and median*10000.

For the broadest distribution (the lower graphs of Figures 20 – 23) most of the median limits and some of the fixed limits are probably too low. The distributions of pounds per application (Figure 23) are generally broader, probably because of the variable meaning of one application. For example, the range of values of pounds per application for indoxacarb covers 8 orders of magnitude. Here, even a limit 10,000 times greater than the median is probably too low.

Although the procedures for determining AI groups in multimodal distributions usually work well, in

some cases the different modes could not clearly be distinguished by either site treated or product used, such as with the bifenthrin (Figure 21) and bromadiolone (Figure 22), each with two modes.

Table 6: *A description of the vertical lines used in the frequency distributions of log of rates of use. These represent the different outlier limits from different criteria. However, two different sets of median criteria are displayed: one for agricultural records with units in acres (shown in the first three lines) and another for all other cases (shown in the next three lines).*

Graph Line	Abbreviation	Description
	med50	Median log (rate times 50) (ag-acres)
	med100	Median log (rate times 100) (ag-acres)
	med200	Median log (rate times 200) (ag-acres)
	med200	Median log (rate times 200) (other cases)
	med5000	Median log (rate times 5000) (other cases)
	med10000	Median log (rate times 10000) (other cases)
	fixed1	Low fixed limit
	fixed2	Medium fixed limit
	fixed3	High fixed limit
	mean5SD	Trimmed mean (log rate) + 5* standard deviation
	mean8SD	Trimmed mean (log rate) + 8* standard deviation
	mean10SD	Trimmed mean (log rate) + 10* standard deviation
	mean12SD	Trimmed mean (log rate) + 15* standard deviation
	median_rate	Median log rate
	pct99	99th percentile of log rates
	pct99_9	99.9th percentile of log rates

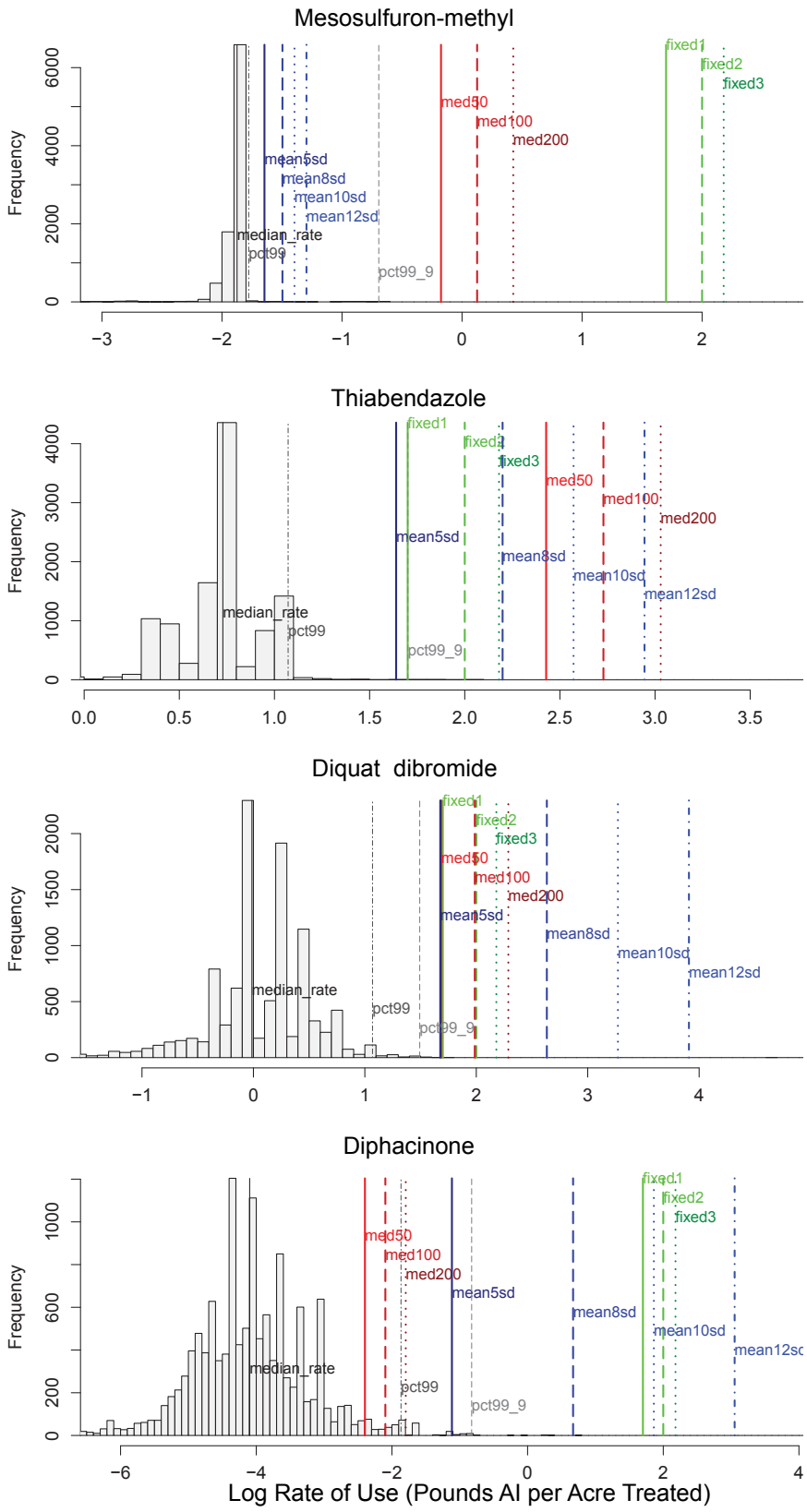

Figure 20: *Frequency distribution of log rate of use for mesosulfuron-methyl, thiabendazole, diquat dibromide, and diphacinone for agricultural records with units treated in acres.*

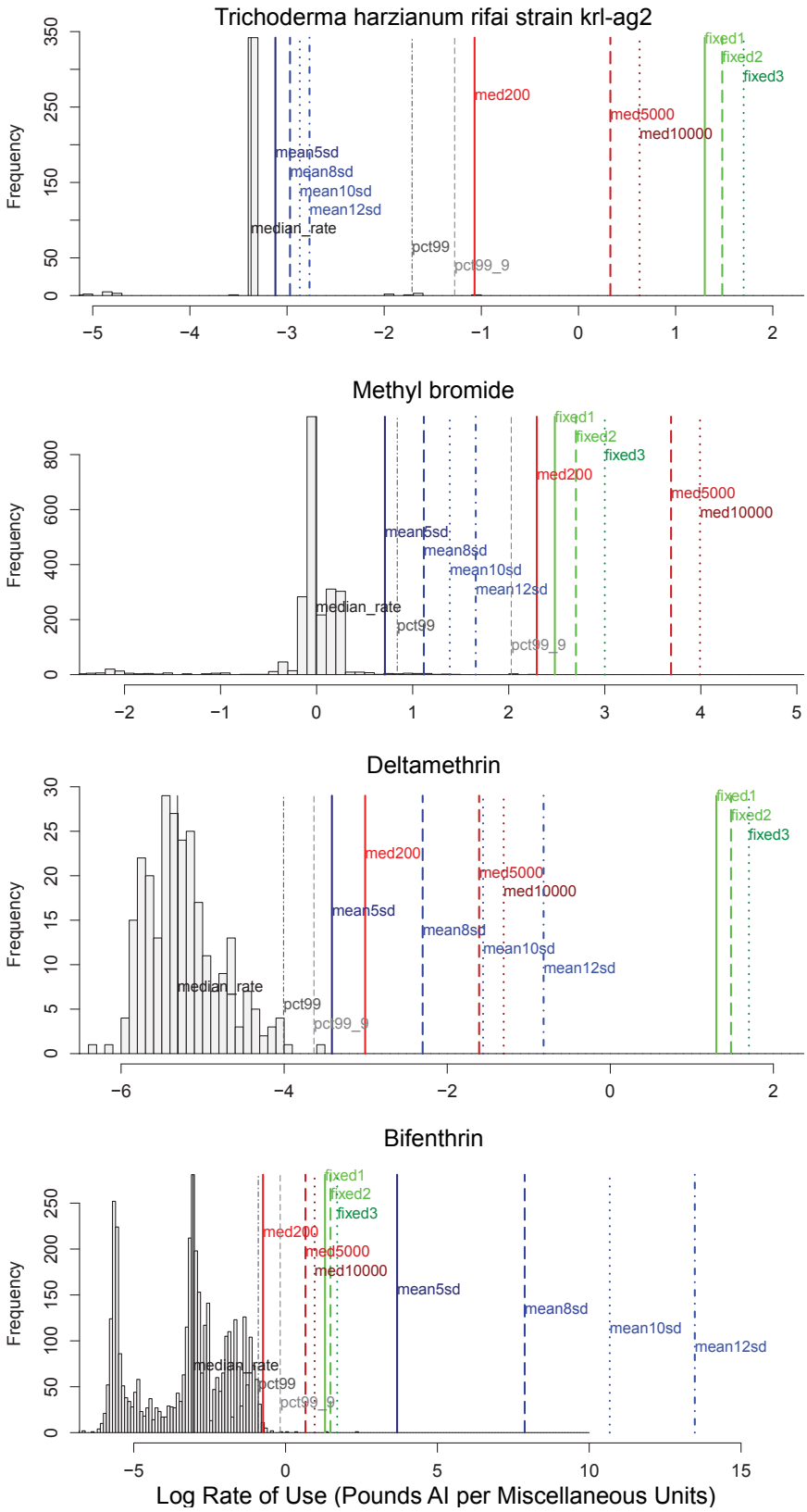

Figure 21: *Frequency distribution of log rate of use for trichoderma harzianum rifai strain krl-ag2, methyl bromide, deltamethrin, and bifenthrin for agricultural records with units treated in miscellaneous units.*

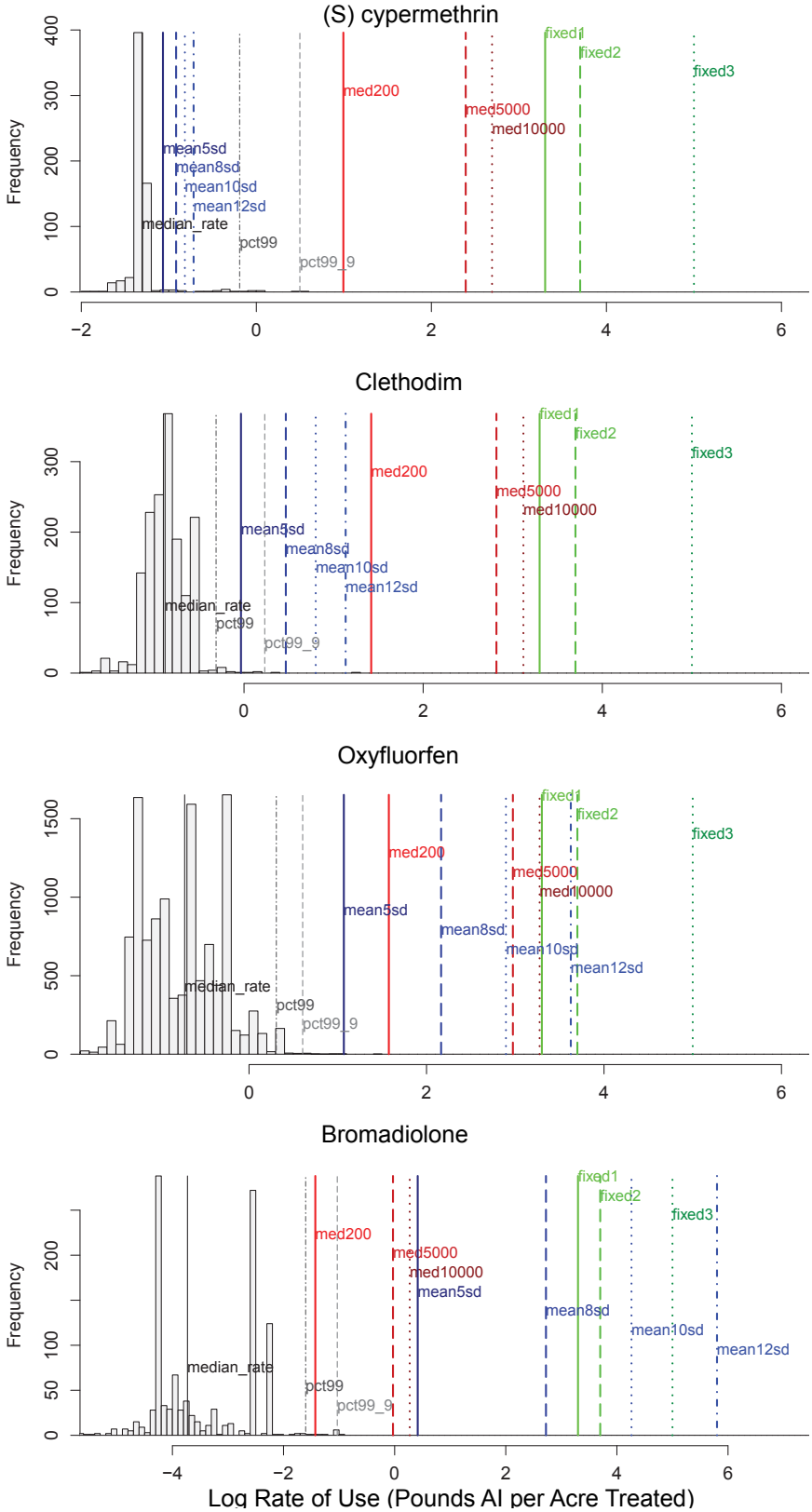

Figure 22: *Frequency distribution of log rate of use for (s)-cypermethrin, clethodim, oxyfluorfen, and bromadiolone for non-agricultural records with units treated in acres.*

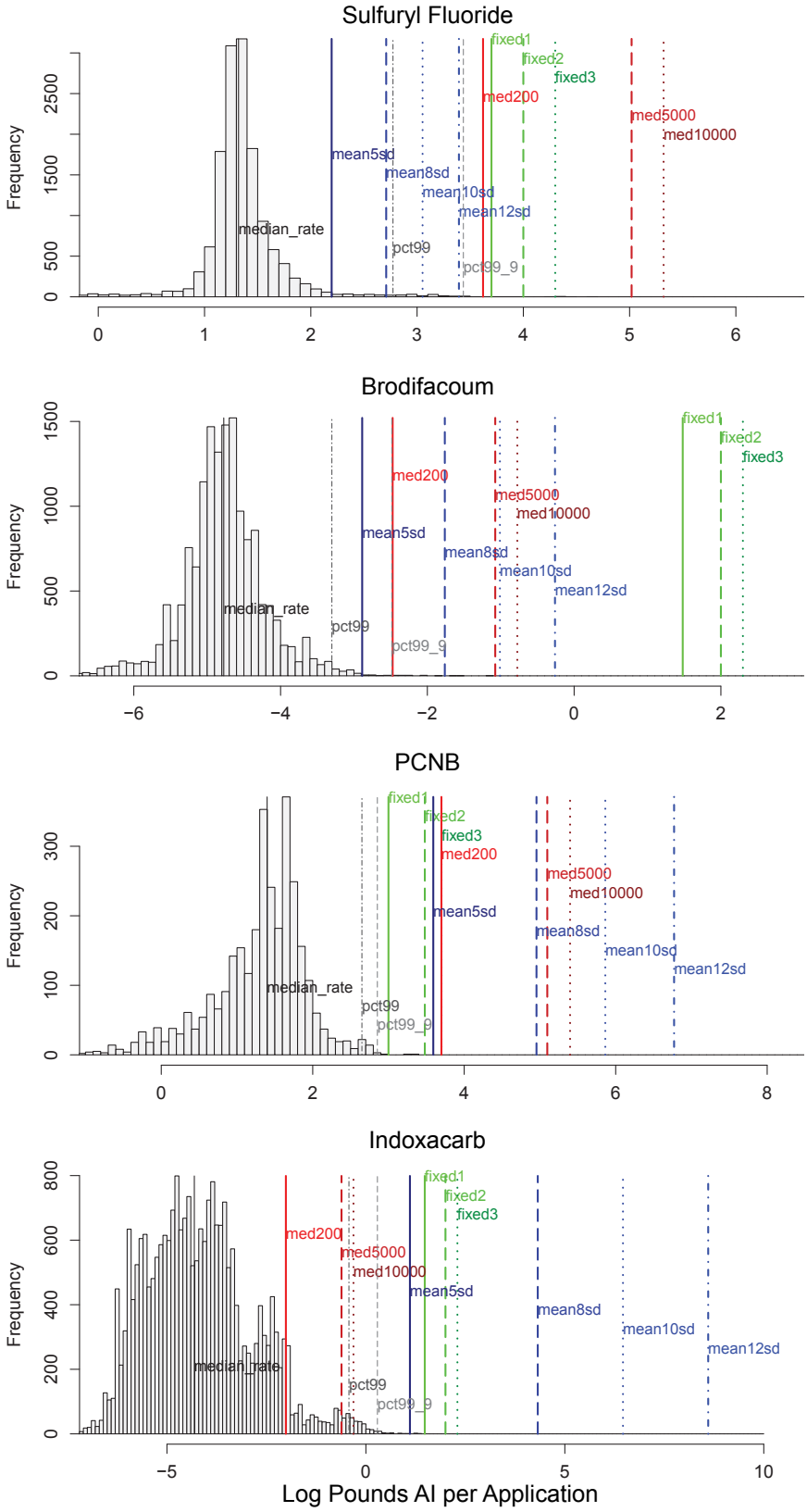

Figure 23: *Frequency distribution of log pounds of AI per application for sulfuryl fluoride, brodifacoum, PCNB, and indoxacarb for non-agricultural records.*

6.2 Number of records identified as outliers by each criterion

Another way to evaluate the different outlier criteria is to compare the percentage of PUR records that are identified as outliers by each of the criteria (Table 7 – Table 11). Currently, about 0.25% of the PUR records are flagged by the current median*50 criterion (Table 11); these are reported as possible errors and sent to the counties for correction. Any new criteria used to determine which records are reviewed by the counties needs to flag about the same number of records as currently since the counties cannot process more errors. More stringent criteria can still be adopted to flag additional records, which can be used by people who need to do more in depth analyses.

6.2.1 Evaluation of fixed limits

The fixed limit criterion uses different outlier limits for different kinds of records (production agriculture or other kind of application), units treated (acres, cubic feet, pounds, or miscellaneous), types of AI (high rate, medium rate, or normal rate AI or adjuvant AI), and sites (water or non-water site, where water site includes the PUR sites "water area", site_code 65000 and "industrial water", site_code 65503) (Tables 2 and 5). Three different fixed limit values were chosen for each kind of record, providing different levels of outlierness. Fixed limits are used because the other criteria depend on the existence of previous records for an AI or AI group, and there may be few such records, or the previous records may have a high percent of unusual reported rates.

The primary method for determining what values to use for these fixed outlier limits used percent of recent PUR records greater than different values (Tables 7 to 10). In these tables the values that are shaded determine the fixed limit; the three levels of shading represent the three different fixed limits and correspond to the limits given in table 2 (fixed1, fixed2, and fixed3). The fixed limits are meant to capture the most extreme rates, so they should flag a small percent of records, something less than 0.25%. Somewhat arbitrally the limits were set in most cases to values where approximately 0.05, 0.03, and 0.02% of the records were greater than the limits. Table 7 shows the results for pounds per unit treated in production agricultural records, table 8 for pounds per unit treated in non-agricultural records, table 9 for pounds per application in non-agricultural records for all sites other than water, and table 10 for pounds per application in water sites.

In Table 7 agricultural records are separated into several groups based on the unit treated, on the AI rate type, and on whether the AI is typically an adjuvant or not. For records with units in acres, the percentages were calculated from the 2013 and 2014 PUR. For normal rate AIs with miscellaneous units, records came from the last 4 years (2011 to 2014). For other units, records came from the last 15 years (2000 to 2014). The number of records in each situation is given in the "Number Records" column.

The decision on how many years of data to use was based on several factors. One factor was the number of records in a set. If there were few records per year, then more years were used in order to have a large set for determining limits. For sets with many records per year, only data from the most recent years were used because typical rates sometimes change over time so the rates in the most recent years should more accurately reflect use now or in the near future. Also, some years, especially 2009 and 2010, had an unusual number of high rates (Figures 24 to 25). Thus for records with normal rate AIs reporting in miscellaneous units, the percentages were calculated on data from years 2011 to 2014 and did not include 2010 even though that would have provided more records. The other columns in the table show the percent of records greater than several different values of rate of use, which are given in the column head. The meaning of these numbers depend on the unit treated shown in the "Unit Treated" column. For example, in the first 4 rows with units in acres, the numbers in

the column head after "Number Records" are pounds of AI per acre, and for the next three rows, the numbers are pounds of AI per cubic feet.

For normal rate agricultural AIs with units in acres, 0.050% of the records during 2013 and 2014 were greater than 50 pounds of AI per acre, 0.031% were greater than 100 pounds per acre, and 0.021% were greater than 150 pounds per acre. These 3 rates were assigned to the 3 fixed outlier limits (fixed1, fixed2, and fixed3, respectively). These limits seem reasonable based on typical rates of use of most AIs. For medium rate agricultural AIs with acres the percent of records greater than 200, 500, and 1000 pounds per acre were 0.050, 0.032, and 0.017%. For high rate agricultural AIs the percent of records greater than 1000, 2000, and 5000 pounds per acre were 0.083, 0.077, and 0.045%. These were higher percentages than with the other AIs, but still less than 0.25%. However, these limits seem reasonable even for the highest rate fumigants which seldom, if ever, have maximum labels rates over 1000 pounds per acre. For adjuvants, the percent of records greater than 50, 100, and 200 pounds per acre were 0.022, 0.013, and 0.009%. The limits were chosen higher for adjuvants since they often do not have maximum label rates, making it difficult to know what rate is too high, and because they may be of less interest.

For agricultural records with units in cubic feet or pounds, it becomes more difficult to decide on reasonable high limits, because there are few such records, even over a 15 year period. For applications where unit treated was cubic feet the highest reported rate over the last 15 years was 6.0 pounds of AI per cubic foot and the next highest rate was 2.4. Setting a fixed limit of 5 pounds per cubic foot would flag 0.21% of the records, but that is just one record over the last 15 years. The other two limits were set to 10 and 15 pounds per cubic foot. A similar reasoning was used for agricultural records with units in pounds. The largest rate over the last 15 years was 127 pounds AI per pound treated, the next largest rate was 63, and the third largest rate was 11. The limits for these kinds of records were set at 20, 50, and 150 pounds per pound.

There are more agricultural records with miscellaneous units ("U") making it easier to determine outlier limits. For normal rate AIs, limits of 20, 30, and 50 flagged 0.049, 0.024, and 0.016% of records, respectively, from the last 4 years. Four years were used because of the unusual number of high rates in previous years and because there were enough records using just 4 years. For medium and high rate AIs, 15 years were used. For medium rate AIs, there were two records with 106 pounds per miscellaneous units, one record with 80 pounds per unit, and many records less than 60 pounds per unit. Limits were set at 100, 200, and 500 pounds per unit, which are all higher than nearly all rates in the last 15 years. For high rate AIs, limits of 300, 500, and 1000 would flag 0.027, 0.027, and 0% of records in the past 15 years.

The same process was carried out for non-agricultural (monthly summary) records (Table 8). For non-agricultural records with no reported units treated, outlier limits are based on pounds of AI per application. For these records, AIs were separated into more groups: non-adjuvant AIs were classified into 5 different groups based on the typical pounds per application used (see Tables 3 and 4). The percentage of records greater than each value in a set of values of pounds per application and the outlier limits chosen are given in Tables 9 and 10.

Table 7: *Percent of PUR production agricultural records of different kinds with rates (pounds per unit treated) greater than each value in a set of values. These percentages were used to determine reasonable fixed outlier limits. The kind of record is determined by unit treated and AI rate type. The units treated are acres ("A"), cubic feet ("C"), pounds ("P"), and miscellaneous (such as bins, boxes, trees, etc.) ("U"). The AI rate types are normal rate, medium rate, high rate, and adjuvant. The percentages are calculated from all PUR records in the latest years, the number of years varying from 2 to 15 years (given in the "Num Years" column). Also given is the number of records for each situation. The remaining columns give the percent of records with rates greater than the number in the column head. The cells that are shaded are the percentages chosen for the fixed outlier limits, with the level of shading representing three different limits.*

Unit	AI Rate	Num	Number	Percent of PUR Records Greater Than Each Rate in Pounds/Unit Treated																				
Treated	Type	Years	Records	1	5	10	15	20	30	40	50	70	100	120	150	200	300	500	1000	1500	2000	5000	10K	50K
A	Normal	2	4,269,341	19.75	1.82	0.50	0.25	0.17	0.11	0.07	0.05	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.00
A	Medium	2	483,954	92.43	50.51	30.10	15.06	11.23	6.03	3.52	1.31	0.58	0.18	0.11	0.07	0.05	0.04	0.03	0.02	0.01	0.01	0.01	0.01	0.01
A	High	2	15,658	98.59	97.13	96.88	96.77	96.67	96.56	96.35	95.52	94.67	74.95	71.11	63.53	31.89	10.49	0.11	0.08	0.08	0.08	0.04	0.04	
A	Adjuvant	2	1,164,836	6.39	0.20	0.10	0.07	0.05	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00
C	Normal	15	486	0.82	0.21																			
C	Medium	15	2																					
C	High	15	14																					
P	Normal	15	324	2.16	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.31	0.31	0.31										
P	Medium	15	6	50.00	33.33	16.67																		
P	High	15	5																					
U	Normal	4	12,345	2.43	0.22	0.10	0.05	0.05	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01							
U	Medium	15	2,729	7.73	3.30	1.91	1.25	0.99	0.59	0.44	0.40	0.15	0.07											
U	High	15	3,646	36.12	1.67	1.12	0.91	0.88	0.77	0.71	0.71	0.69	0.44	0.25	0.22	0.16	0.03	0.03						

Table 8: *Percent of PUR non-agricultural records of different kinds with rates (pounds per unit treated) greater than each value in a set of values. These percentages were used to determine reasonable fixed outlier limits. The kind of record is determined by the unit treated and AI rate or site type. The units treated are acres ("A"), cubic feet ("C"), pounds ("P"), and miscellaneous (such as bins, boxes, trees, etc.) ("U"). The AI rate types are normal rate, medium rate, and high rate. The site type is water body or other site. The percentages are calculated from all PUR records in the latest years, the number of years varying from 2 to 15 years (given in the "Num Years" column). Also given is the number of records for each situation. The remaining columns give the percent of records with rates greater than the number in the column head. The cells that are shaded are the percentages chosen for the fixed outlier limits, with the level of shading representing three different limits.*

Unit	AI Rate/Site	Num	Number	Percent of PUR Records Greater Than Each Rate in Pounds/Unit Treated													
Treated	Type	Years	Records	1	5	10	15	20	30	40	50	70	100	150	200	300	500
A	Normal	4	203,123	39.38	5.67	1.74	1.05	0.77	0.55	0.47	0.42	0.35	0.27	0.22	0.21	0.13	0.08
A	Medium	4	2,312	61.07	31.92	18.51	12.07	7.87	4.97	3.20	2.12	1.38	0.61	0.35	0.30	0.26	0.09
A	High	4	4,724	99.11	98.71	98.29	98.07	97.99	97.73	97.21	95.87	93.69	90.43	85.41	75.38	48.31	0.55
A	Water Site	4	7,603	73.54	29.84	16.74	12.23	9.69	7.34	6.26	4.84	4.01	2.99	2.09	1.50	1.05	0.63
C	Normal	15	18,821	0.61	0.16	0.11	0.06	0.05	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.01	0.01
C	Medium	15	4														
C	High	15	2,955	1.02	0.37	0.14	0.07	0.07									
C	Water Site	15	692	22.25	6.07	2.89	2.02	1.73	1.30	1.16	1.01	1.01	0.87	0.87	0.87	0.72	0.72
P	Normal	15	41,632	0.14	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.00	0.00	0.00	0.00
P	Medium	15	101	0.99													
P	High	15	2,571	0.74	0.04	0.04	0.04	0.04									
P	Water Site	15	389	0.26													
U	Normal	15	27,656	29.36	16.72	12.17	9.89	8.76	7.30	6.58	6.11	5.29	3.75	1.49	1.01	0.68	0.38
U	Medium	15	129	11.63	8.53	6.98	3.10	2.33	1.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U	High	15	1,406	8.04	4.55	3.27	2.49	1.64	1.42	1.21	1.07	1.00	0.57	0.57	0.50	0.28	0.14
U	Water Site	15	12,182	78.14	56.97	48.61	41.56	36.95	32.13	29.35	23.67	19.33	15.99	12.70	11.11	8.64	7.44

Unit	AI Rate/Site	Num	Number	Percent of PUR Records Greater Than Each Rate in Pounds/Unit Treated													
Treated	Type	Years	Records	1000	2000	3000	4000	5000	10K	20K	30K	50K	70K	100K	200K	300K	500K
A	Normal	4	203,123	0.06	0.05	0.05	0.04	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
A	Medium	4	2,312	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09			
A	High	4	4,724	0.08	0.06	0.04	0.02										
A	Water Site	4	7,603	0.26	0.11	0.11	0.11	0.07	0.03	0.03	0.03	0.03	0.03	0.01	0.01	0.01	0.01
C	Normal	15	18,821														
C	Medium	15	4														
C	High	15	2,955														
C	Water Site	15	692	0.43	0.29	0.14	0.14										
P	Normal	15	41,632														
P	Medium	15	101														
P	High	15	2,571														
P	Water Site	15	389														
U	Normal	15	27,656	0.22	0.13	0.09	0.06	0.05	0.04	0.04	0.02	0.01	0.00				
U	Medium	15	129	0.00	0.00	0.00	0.00	0.00									
U	High	15	1,406	0.07	0.07	0.07	0.07	0.07	0.07	0.07							
U	Water Site	15	12,182	5.21	3.38	2.54	2.01	1.64	1.12	0.71	0.57	0.19	0.05	0.02	0.01	0.01	

Table 9: *Percent of PUR non-agricultural records on non-water sites of different AI rate types with pounds of AI per application greater than each value in a set of values. These percentages were used to determine reasonable fixed outlier limits. The kind of record is determined by the AI rate type, which include adjuvants and five groups of AIs from those used at the highest amounts ("High Lbs AI1") to the lowest amounts ("High Lbs AI5"). The percentages are calculated from all PUR records in the latest years, the number of years varying from 2 to 15 years (given in the "Num Years" column). Also given is the number of records for each situation. The remaining columns give the percent of records with rates greater than the number in the column head. The cells that are shaded are the percentages chosen for the fixed outlier limits, with the level of shading representing three different limits.*

AI Rate	Num	Number	Percent of PUR Records Greater Than Each Pounds per Application													
Type	Years	Records	30	50	100	150	200	500	1000	2000	3000	4000	5000	6000	7000	8000
High Lbs AI5	2	878,430	0.05	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.00					
High Lbs AI4	2	358,315	4.00	2.32	1.20	0.79	0.56	0.18	0.07	0.02	0.02	0.01	0.01	0.00	0.00	0.00
High Lbs AI3	15	151,906	31.41	22.22	14.97	11.65	9.84	5.72	3.12	1.31	0.53	0.19	0.09	0.05	0.03	0.03
High Lbs AI2	15	13,037	61.26	49.01	37.65	30.17	24.55	12.57	7.39	4.23	2.97	2.20	1.79	1.37	1.25	1.10
High Lbs AI1	15	472	63.56	54.66	48.52	45.76	34.96	26.91	12.71	4.45	3.81	3.81	3.60	3.39	3.39	
Adjuvants	15	199,837	4.48	2.60	1.17	0.70	0.47	0.17	0.05	0.02	0.01	0.01	0.01	0.01	0.01	0.00

AI Rate	Num	Number	Percent of PUR Records Greater Than Each Pounds per Application														
Type	Years	Records	9000	10K	15K	20K	30K	40K	50K	60K	70K	80K	90K	100K	200K	300K	400K
High Lbs AI5	2	878,430															
High Lbs AI4	2	358,315	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00							
High Lbs AI3	15	151,906	0.03	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
High Lbs AI2	15	13,037	0.92	0.91	0.64	0.26	0.13	0.04									
High Lbs AI1	15	472	2.97	2.75	2.12	1.06	1.06	1.06	0.85	0.42	0.42	0.42	0.42	0.21	0.21	0.21	
Adjuvants	15	199,837	0.00	0.00													

Table 10: *Percent of PUR non-agricultural records on water sites of different AI rate types with pounds of AI per application greater than each value in a set of values. These percentages were used to determine reasonable fixed outlier limits. The kind of record is determined by the AI rate type, which include adjuvants and five groups of AIs from those used at the highest amounts ("High Lbs AI1") to the lowest amounts ("High Lbs AI5"). The percentages are calculated from all PUR records in the latest 15 years. Also given is the number of records for each situation. The remaining columns give the percent of records with rates greater than the number in the column head. The cells that are shaded are the percentages chosen for the fixed outlier limits, with the level of shading representing three different limits.*

AI Rate	Num	Number	Percent of PUR Records Greater Than Each Pounds per Application													
Type	Years	Records	100	150	200	500	1000	1500	2000	3000	4000	5000	6000	7000	8000	9000
High Lbs AI5	15	903	7.97	0.44	0.33	0.22	0.11									
High Lbs AI4	15	17,222	4.42	2.84	2.11	0.70	0.29	0.23	0.16	0.11	0.10	0.04	0.03	0.03	0.03	0.02
High Lbs AI3	15	6,570	21.39	16.85	15.22	8.71	4.19	2.39	1.83	1.05	0.37	0.27	0.23	0.20	0.14	0.12
High Lbs AI2	15	6,680	30.21	26.68	22.13	14.22	9.39	6.38	5.07	3.50	2.38	1.98	1.71	1.63	1.53	1.53
High Lbs AI1	15	186	90.86	90.86	90.32	89.25	89.25	89.25	83.87	83.33	80.11	79.03	73.66	71.51	61.29	60.22
Adjuvants	15	753	15.14	14.08	13.68	11.95	11.02	8.76	8.76	5.84	5.71	4.38	2.12	0.93	0.66	0.40

AI Rate	Num	Number	Percent of PUR Records Greater Than Each Pounds per Application														
Type	Years	Records	10K	15K	20K	30K	40K	50K	60K	70K	80K	90K	100K	150K	200K	250K	500K
High Lbs AI5	15	903															
High Lbs AI4	15	17,222	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01				
High Lbs AI3	15	6,570	0.11	0.08	0.06	0.03	0.02	0.02									
High Lbs AI2	15	6,680	1.45	1.38	0.48	0.33	0.06	0.01	0.01								
High Lbs AI1	15	186	46.24	30.11	20.43	11.29	8.06	5.91	5.38	2.69	1.61	1.08	0.54	0.54			
Adjuvants	15	753	0.27														

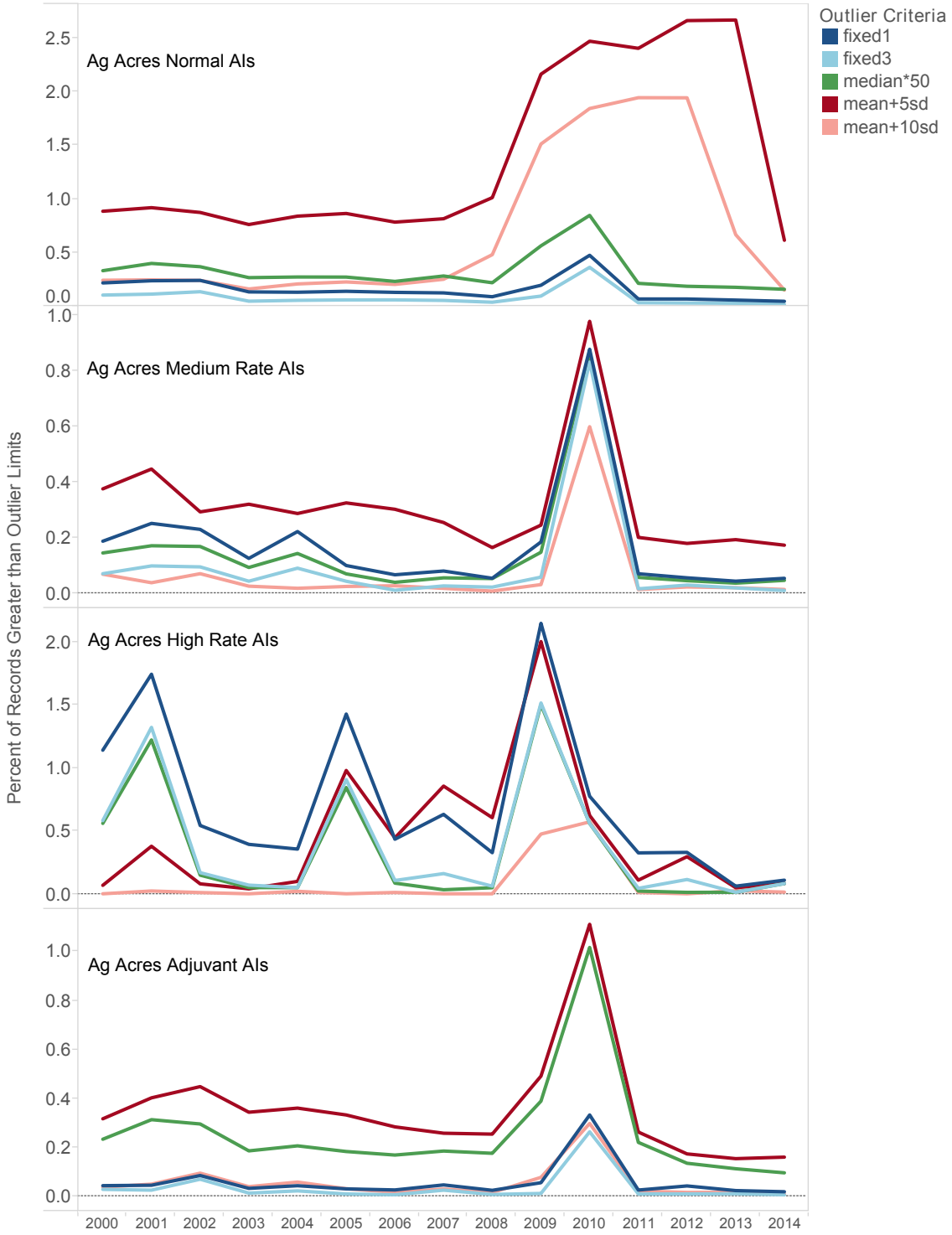

Figure 24: *Percent of records in the PUR with rates of use greater than each of four outlier criteria by year for production agriculture application with unit treated in acres.*

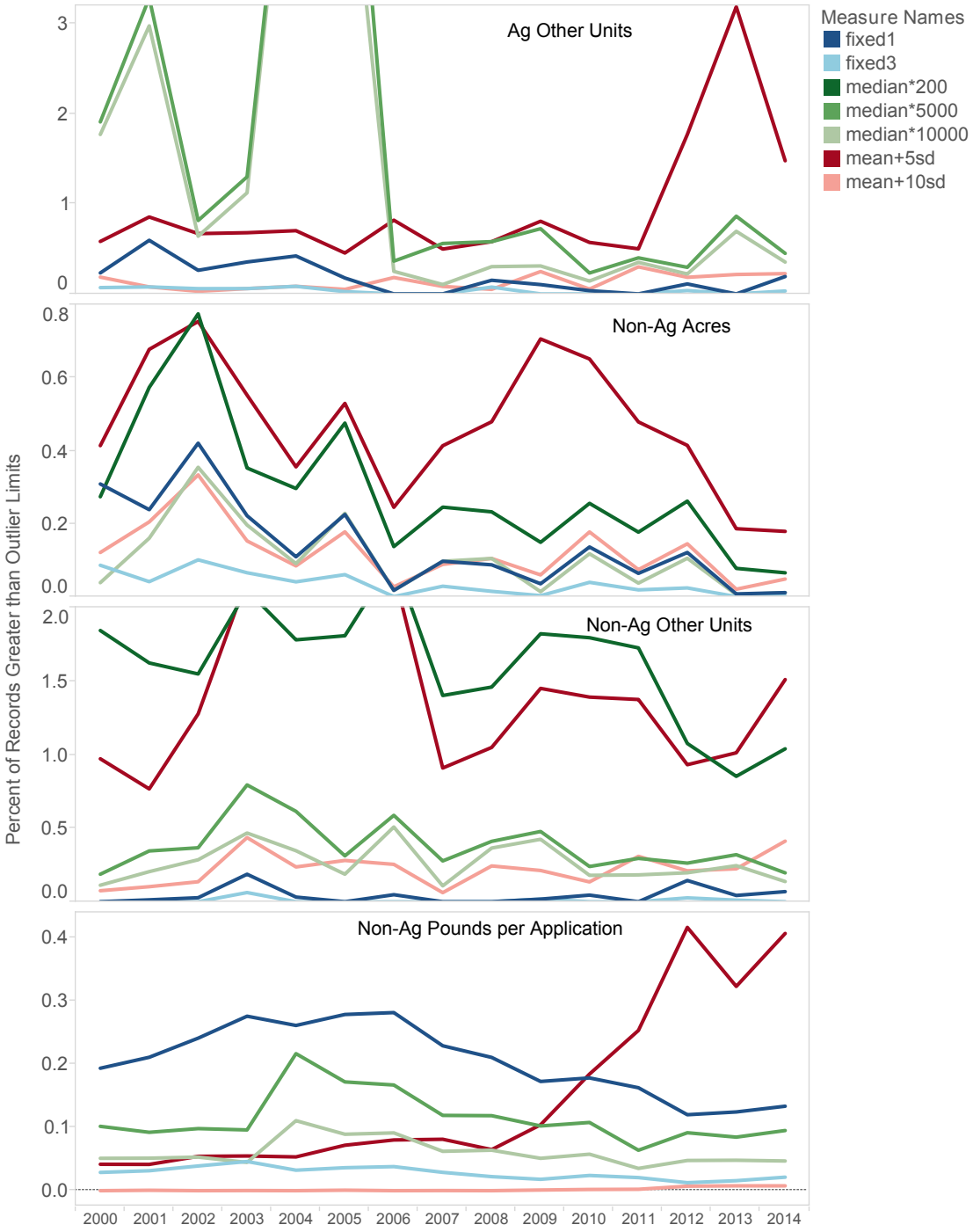

Figure 25: *Percent of records in the PUR with rates of use greater than each of four outlier criteria by year for production agriculture application with unit treated other than acres and for non-agricultural records.*

6.2.2 Evaluation of all criteria

Table 11 gives the medians of yearly percentages of rates greater than each criterion limit for years 2000 to 2014 for several different types of uses. Yearly medians were used here rather the percentages of all rates from recent years, which was used in the fixed outlier tables (Tables 7 to 10), because in some years the percentages were very different than in most other years (Figures 24 and 25). No reason could be found for the unusually high rates in 2009 or 2010 but they don't seem to be representive of most years. The presence of these high values results in an unreasonably high rate for total percentage. The median of the yearly percentages is not affected by a few large values.

The percent of production agricultural records with unit treated in acres that were greater than some criterion limit varied from 0% to 2.21% (first 4 rows in Table 11). Excluding the criterion mean+3*SD, the highest percent was 0.88%. Most of the percentages were less than 0.25%, which is the percent of records that is currently sent to the counties to be checked. Most of the highlighted values were close to 0.25% and indicate suggested limits to use for sending records back to the counties. The largest percentages in most cases were for normal rate AIs (the first row in Table 11), which may not be surprising since this group contains the most widely varying rates. The percentages chosen for fixed limits were less than 0.25% because the fixed limits are meant to find only the most extreme values. In Table 7 the three fixed limits for the normal and medium rate AIs with unit treated in acres were chosen at values where 0.05, 0.03, and 0.02% of the records were flagged, which are lower percentages than the ones in table 11. This is because in the first table the percentages were calculated for the most recent two years and in the second table all years were used, and the more recent years had fewer high values.

The percent of agricultural records greater than each criterion limit with other units treated (cubic feet, pounds, and miscellaneous) varied from 0.4 to 8.9% (the fifth row in Table 11). The percentages for the fixed and trimmed mean criteria were reasonably close to the percentages for records with acres treated. However, the percentages for the median criteria were much larger. This is not surprising because the variances of rates of use in other units are expected to be larger and the median criteria do not make use of the variance, while the trimmed means do. The fixed limits also do not consider the variance but these limits were set higher for these other units.

The percentages for non-agricultural records with acres treated was similar to the agricultural records with acres treated (the sixth row in Table 11), except that the median percentages were larger, again probably due to larger variances in rates of use. For the non-agricultural records with other units, fixed limit percentages were low, and the median and trimmed mean percentages were high. For the non-agricultural records using pounds of AI per application, the fixed limits are low, the median percentages are high, and the trimmed mean percentages are very low.

The percentages for the new median*50 were close to the current median*50, which is reasonable since they are very similar criteria. However, the percentages for the new median*50 were slightly higher in most cases than the current version. This is because the current version misses some outliers for situations with few records for a particular pesticide product and site treated.

Table 11: *Percent of different kinds of PUR records with rates of use greater than each of several outlier criteria. The values are medians of yearly percentages from years 2000 to 2014. The kind of record is determined by record type, unit treated, and AI rate type. The record types are production agricultural ("Ag") and all other types of applications, mostly non-agricultural ("Non-Ag"). The units treated are acres and all other units, which includes cubic feet, pounds, and miscellaneous (such as bins, boxes, trees, etc.). The AI rate types are normal rate, medium rate, high rate, and adjuvant. In all rows, except the last row, the measure of rate is pounds of AI per unit treated; in the last row the rate is pounds per application (unit treated is listed as "Appl"). "Number of Records" are total number of records for each set of records. The shaded numbers indicate the recommended outlier criteria.*

				Outlier Criteria																			
Record Type	Unit Treated	AI Rate Type	Number Records	FIX1	FIX2	FIX3	MED *50	MED *100	MED *200	MED *500	MED *1000	MED *5000	MED *10K	MED *50K	MEAN +3*SD	MEAN +5*SD	MEAN +7*SD	MEAN +8*SD	MEAN +10*SD	MEAN +12*SD	MEAN +15*SD	OLD FIX	OLD MED50
Ag	Acre	Normal	25,152,125	0.13	0.07	0.05	0.27	0.16	0.07	0.04	0.04	0.03	0.02	0.01	2.21	0.88	0.46	0.36	0.24	0.18	0.10	0.05	0.25
Ag	Acre	Medium	3,162,921	0.10	0.06	0.05	0.07	0.05	0.04	0.02	0.02	0.01	0.01	0.01	0.73	0.29	0.10	0.07	0.03	0.02	0.01	0.10	0.13
Ag	Acre	High	134,285	0.43	0.29	0.11	0.08	0.05	0.04	0.02	0.02				2.12	0.29	0.03	0.02	0.01			0.43	0.94
Ag	Acre	Adjuvant	6,531,081	0.04	0.02	0.01	0.19	0.12	0.07	0.02	0.02	0.01	0.01	0.01	1.94	0.32	0.12	0.08	0.03	0.02	0.01	0.01	0.18
Ag	Other		60,173	0.15	0.06	0.03	8.92	6.88	3.88	2.04	1.39	0.72	0.35	0.08	2.54	0.68	0.28	0.21	0.08	0.06	0.02		4.66
Non-Ag	Acre		592,040	0.11	0.09	0.03	0.64	0.35	0.26	0.19	0.14	0.10	0.10	0.04	1.62	0.48	0.22	0.15	0.10	0.07	0.03	0.44	0.44
Non-Ag	Other		108,559	0.03	0.01		3.07	2.25	1.72	1.09	0.69	0.32	0.20	0.09	2.94	1.37	0.71	0.39	0.22	0.12	0.04		1.32
Non-Ag	Appl		8,177,976	0.21	0.06	0.03	5.47	3.22	1.78	0.80	0.44	0.10	0.05	0.01	1.52	0.08	0.01						

6.3 Relative ranking of outlier limits

The graphs of distributions of rates overlaid with lines for outlier limits (Figures 20 – 23) illustrated that sometimes a criterion was small relative to other criteria and sometimes large. Ideally, the best criterion should be less extreme relative to other criteria. Thus, another method to evaluate the different criteria would be to examine the relative ranking of each criterion. For production agricultural records with unit treated in acres and normal rate AIs with more than 5 records, comparing criteria median*50, median*100, median*200, mean+5*SD, mean+8*SD, mean+10*SD, mean+12*SD, fixed1, fixed2, and fixed3, the criterion mean+5*SD was most often the smallest value (about 15% of the time) (Figure 26). The criterion median*50 was the smallest value about 10% of the time and fixed1 about 2.9% of the time. On the other hand, the criterion mean+12*SD was the largest value 53% of the time, the fixed3 criterion about 34% of the time, and median*200 about 12% of the time. Criteria that were rarely the smallest or largest include median*100, mean+8*SD, mean+10*SD, and fixed2. Based on this graph, none of these four criteria were obviously better than the others. However, in a previous analysis, which used only median*50 and simpler fixed limits, the trimmed mean criteria were clearly better. The median and fixed limits used now are more sophisticated and take into account different kinds of AI, records, and units resulting in better results for these criteria.

It was somewhat surprising to see that the trimmed mean limits were often among the largest values. From the frequencey distribution graphs with outlier limits (Figures 20 – 23) it is clear that for narrow distributions, the trimmed means do better and as distributions become broader, the trimmed mean limits become larger relative to the median limits. The fact that trimmed mean limits were often large relative to fixed and median limits reflects the fact that there were more broad distributions than narrow ones.

The relative ranking graphs for the other kinds of AIs in production agriculture with unit treated in acres (the lower three graphs in Figure 26) were fairly similar to that for normal AIs. However, for medium and high rate AIs, all the trimmed mean criteria were more often the largest limits; the median criteria were also somewhat more often larger; but the fixed limits were seldom among the larger limits. The adjuvant graph was closer to that of normal AIs except for mean+12*SD, which was more often larger.

The graphs for agricultural records with other units (mostly miscellaneous units) and non-agricultural records with acres and with other units were similar to that for normal AIs (upper three graphs in Figure 27). However, the graph for non-agricultural records for pounds of AI per application was rather different, mainly because the percentiles were more often the largest. However, if the percentiles were removed, the graph for the three criteria with pounds per application would actually be very similar to the graphs for pounds per unit treated.

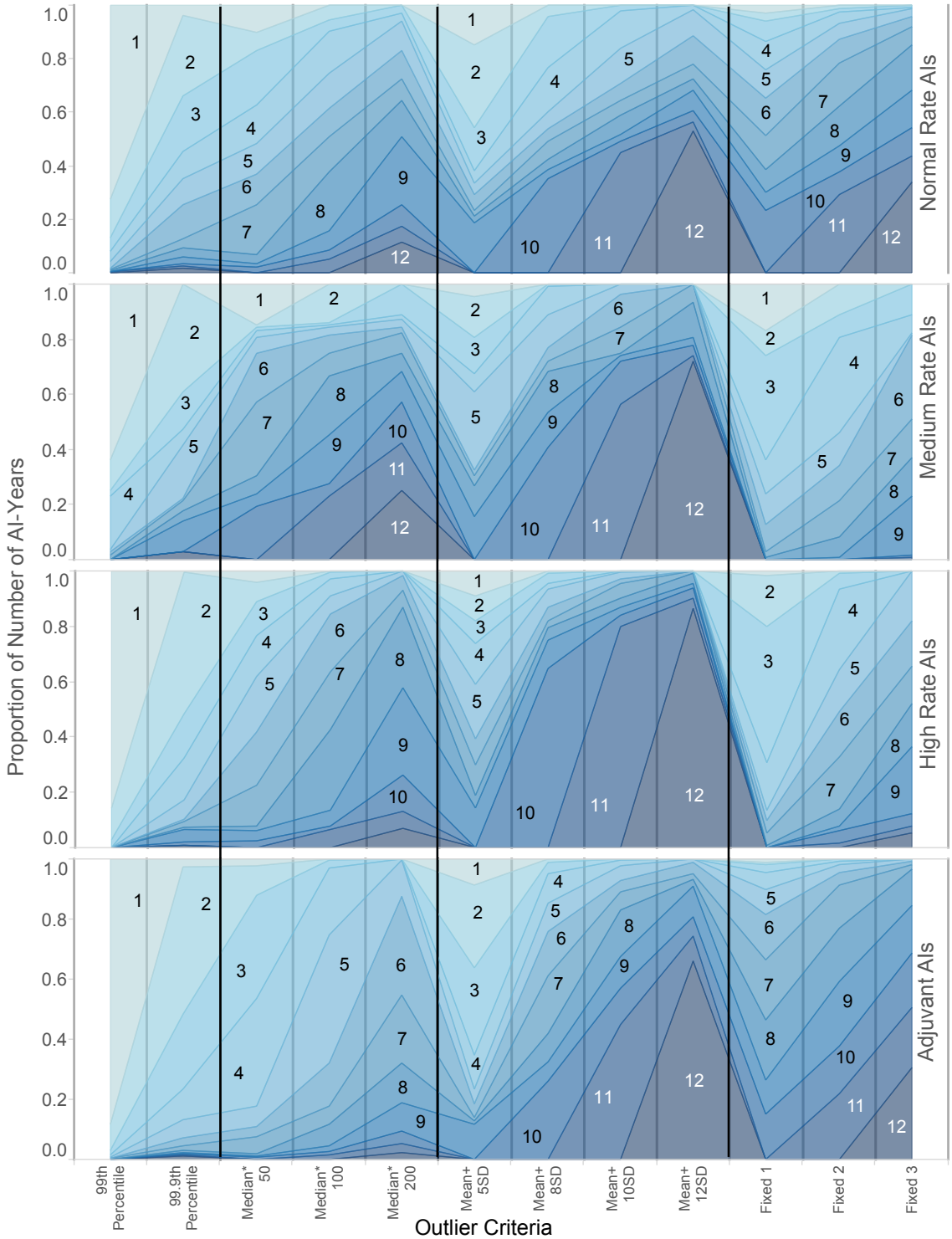

Figure 26: *Relative ranking of the different outlier limits for agricultural records with unit treated in acres. Each outlier criterion is displayed in the vertical strips. The ranks of each criterion are shown both by the shading and the number in each shaded section. For example, the fixed3 criterion for normal rate AIs, on the right side of the top graph was the highest value (rank 12) 34% of the time and was never the lowest value (rank 1); it has the 5th lowest value 2.9% of the time.*

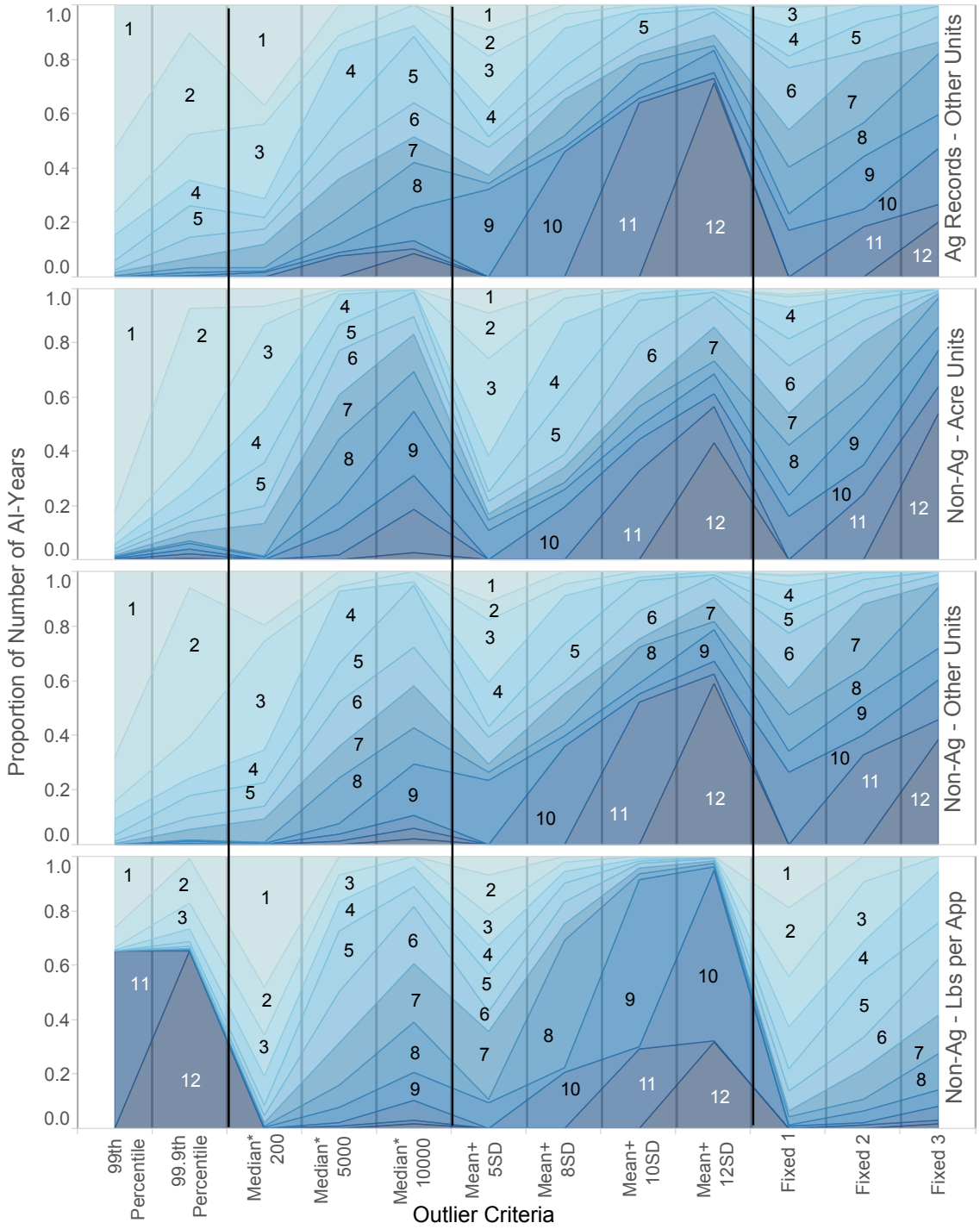

Figure 27: *Relative ranking of the different outlier limits for agricultural records with units treated other than acres and for non-agricultural records. Each outlier criterion is displayed in the vertical strips. The ranks of each criterion are shown both by the shading and the number in each shaded section.*

6.4 Applying criteria to data with known maximum label rates

The final method for evaluating the different outlier criteria was to compare them with the maximum pesticide label rates. DPR has maximum label rates for some of the most commonly applied pesticide products. However, these data have some problems. First, the label rates for a pesticide product often differ considerably depending on what crop or site is treated, but DPR's database has only the maximum label rate among all crops listed on the label. This rate may be too high for some crops. Also, some labels are difficult to interpret and conversions were often needed, so errors are likely to exist. Also, data exist almost exclusively for rates in pounds per acre treated, so the analysis here is only for agricultural records with unit treated in acres.

The relationships between the maximum label rate and the criteria limits for all the products in DPR's maximum label rate table is displayed in Figure 28. Each point in this graph represents the maximum criterion limit and maximum label rate for one product and criterion. As mentioned above, the label rates often differ for different crops and a product may have different criteria limits for different crops. To make this comparison only the maximum criterion limit among all crops for each product was used. The lines are regression lines fitted to these points for each criterion.

Note that the criteria limits are not meant to estimate the maximum label rate but to estimate a rate so extreme it is likely an error. Some growers will sometimes make applications at rates greater than the maximum label rate; these rates are not errors. Therefore, we expect the criteria limits to be greater than the maximum label rates. The solid black like in Figure 28 represents points where the label rates equal the criteria limits. We would expect the points in the graph to lie above this line and most of them do.

Since most of the AIs in DPR's maximum label rate table are normal rate AIs, each fixed limit is the same for all products so the fixed limit lines (the green lines) in Figure 28 are horizontal. For products with low label rates the fixed limits are high and for products with high label rates the fixed limits are too low. The median criteria lines are closest to the black reference line, nearly all slightly greater than the reference line, as expected. Their slope is greater than the reference line slope, meaning that the median limits are greater for products with larger label rates. However, the median*50 line is slightly below the reference line for low rate products, meaning that the median*50 criteria may be too low for these products. The trimmed mean lines all have positive slopes and were all above the reference line, but their slopes were less than that of the reference line, meaning that the trimmed mean limits were larger for products with low label rates.

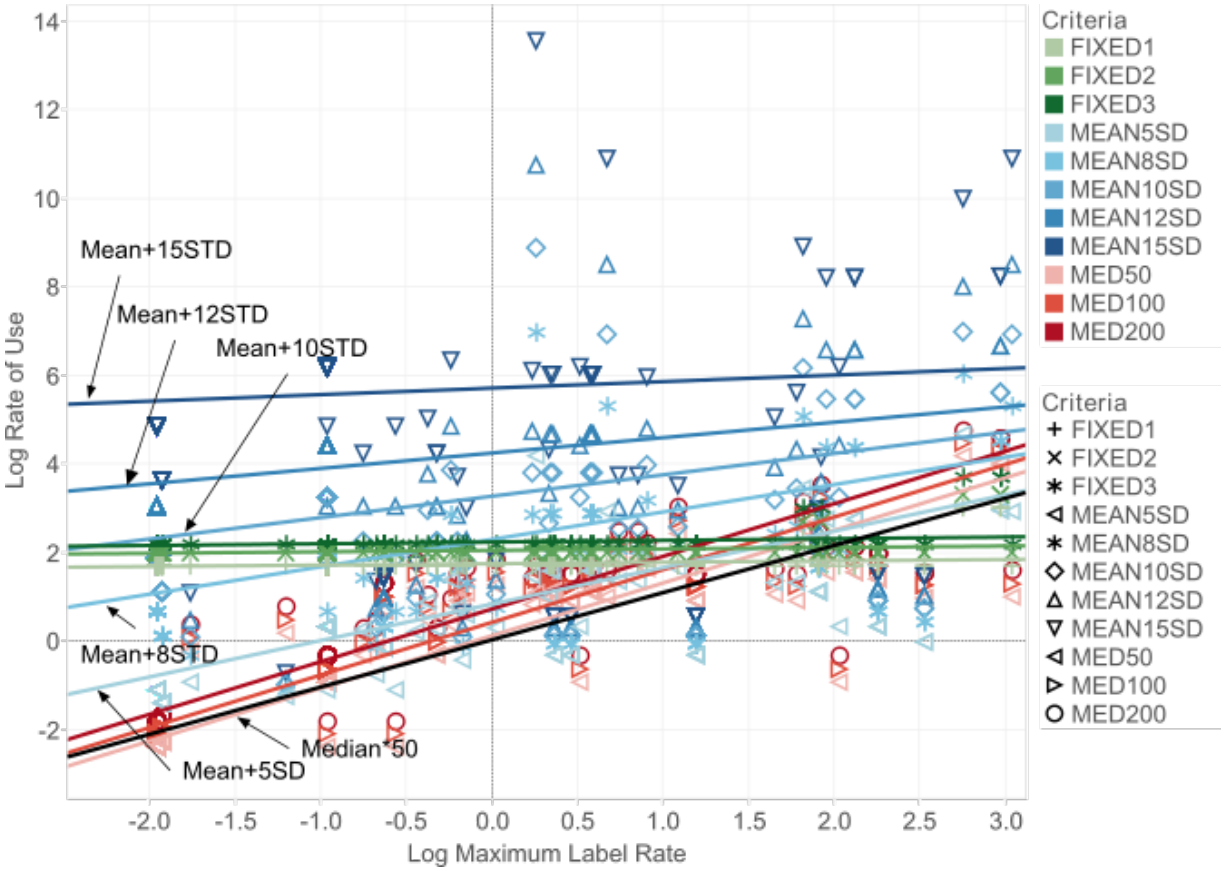

Figure 28: *Comparison of the outlier limits from each criterion to maximum label rates. Each point in this graph represents the maximum criterion limit and maximum label rate for one product and criterion.*

7 Conclusions

Which criteria should be used to flag records with unusually high rates of use? The current system flags records with rates above either the median*50 limit or a fixed limit (using the current definition of these criteria). Because each kind of criterion has advantages and disadvantages, something similar should be adopted for a new system, such as flagging records with rates greater than a fixed, a median, or a trimmed mean limit. However, the graphs of frequency distributions demonstrate that for very broad distributions, the median limits are too low and, therefore, a poor choice. Nearly all distributions where median limits are less than trimmed mean limits have broad distributions, which means the median limits should not be used in these situations. For distributions where a median limit is greater than a trimmed mean limit, any rate flagged by the median criterion will also be flagged by the trimmed mean criterion. Thus, there is no reason to ever use the median rate limits.

The trimmed mean limits appeared to work best in most situations. However, they did not necessarily work well where there were few previous records or where all the previous records happened to have unusually high or low rates. In these situations the fixed limits were probably more reliable. Because the fixed limits were set at very high values, rates above the fixed limits should probably all be flagged.

The conclusion from these two arguments are that rates greater than one of the fixed limits or one of the trimmed mean limits should be flagged. The question is which of the specific fixed and trimmed mean criteria should be used? Table 12 summarizes the conclusions from the previous sections of this report. In most situations the fixed2 limit appeared to work best among the three fixed limits, although on the log scale there usually was not a big difference between the fixed limits. However, remember that the fixed2 limits values differed for different types of records. The best trimmed mean limit varied between different situations, no one limit doing best in all situations. The mean+8*SD and mean+10*SD limits seemed to be best in most situations.

Probably the best way to decide on which limits to use is to look at the percent of records that would be flagged by different combinations of limits (Table 13). The previous tables showing the percent of records flagged (Tables 7 and 11) were calculated from a subset of records of the appropriate type or unit treated but in Table 13 the percentages were based on the number of all records in the PUR. Table 13 gives the percent of records flagged for several specific limits (fixed1, fixed2, fixed3, mean+5*SD, mean+7*SD, mean+8*SD, mean+10*SD, mean+12*SD) and for a few combinations (or suites) of limits. The suites of criteria used were based on the results from Table 11, and the definitions of six criteria suites is given in Table 14. For example, suite 1 flags records with rates greater than mean+7*SD for agricultural records with normal rate AIs and unit treated in acres, mean+5*SD for agricultural records in acres for other AIs, mean+7*SD for agricultural records with other units treated, mean+7*SD for non-agricultural records with units in acres, mean+10*SD for for non-agricultural records with other units treated, and mean+5*SD for non-agricultural records using pounds of AI per application. Suite 2 is like suite 1 except it uses mean+10*SD instead of mean+7*SD for agricultural records in acres for normal rate AIs. The last three suites are the same as the previous three suites except that they include fixed1, fixed2, and fixed3 limits.

Again, ideally the percent of records flagged should be similar to the percentage flagged in the past, which was approximately 0.25%. If the limits used include both fixed and trimmed mean criteria, then the choices from Table 13 are between the last three columns. The median of yearly percent of records flagged from 2000 to 2014 was 0.47% for suite 4, 0.29% for suite 5, and 0.18% for suite 6. The set of criteria that flagged a percentage closest to 0.25% was suite 5 at 0.29%. This percentage was slightly higher than 0.25%, but since this criterion can be applied to more records than the current criteria, this seems a reasonable percentage to flag. The outlier table in the PUR database can record the results of all these criteria to give a user the option to screen outliers more conservatively or liberally.

For example, a person wanting to investigate less extreme outliers could use suite 4 to find more high rates.

The criteria to use to flag outliers depends the record type, unit treated, AI rate type, and site type. The recommendations for determining which records will trigger a warning in the data entry program or cause the records to be sent back to the county for investigation are given in Tables 15 (for determining outliers in pounds per unit treated) and 16 (for determining outliers in pounds of AI per application). These two tables combine the relevant parts of Tables 2, 5, and 11 using suite 5 for the set of criteria limits to use.

Table 12: *Summary of criteria evaluation, showing which criteria appeared to do best in different situations and by different evaluation methods. The evaluations were: 1) visual inspection of frequency distribution graphs; 2) percent of records flagged by the criteria; and 3) relative rankings of outlier limits. No clear results appeared by comparing the criteria to maximum label rates. The different situations include the rate measure (pounds per unit treated or pounds per application) and record type (agricultural or non-agricultural)*

Rate measure, Record type	Visual inspection	Percent records flagged	Relative ranking
Lbs/acre, Ag	mean8, mean10 med100, med200 fixed2	mean5, mean10 med50 fixed2	mean8 med100 fixed2
Lbs/other unit, Ag	mean8, mean10 med200, med5000 fixed2	mean7 med10000 fixed2	mean8 med100 fixed2
Lbs/acre, Non-Ag	mean8, mean10 med200, med5000 fixed2	mean7 med200 fixed2	mean8 med5000 fixed2
Lbs/other unit, Non-Ag	mean8, mean10 med200, med5000 fixed2	mean10 med10000 fixed2	mean8 median5000 fixed2
Lbs/application	mean8, mean10 med5000 fixed2	mean5 med5000 fixed2	mean8 med5000 fixed3

Table 13: *Percent of records in the PUR with rates of use greater than each of several outlier criteria. The values in the last row are medians of yearly percentages for years 2000 to 2014. The suite of criteria limits used in the last six columns is given in Table 14.*

Year	Records	fixed1	fixed2	fixed3	mean 5SD	mean 7SD	mean 8SD	mean 10SD	mean 12SD	Suite1	Suite2	Suite3	Suite4	Suite5	Suite6
2000	2,648,194	0.15	0.09	0.07	0.59	0.31	0.23	0.15	0.11	0.35	0.23	0.15	0.44	0.26	0.18
2001	2,447,935	0.17	0.10	0.08	0.63	0.31	0.23	0.15	0.11	0.37	0.25	0.15	0.48	0.29	0.19
2002	2,483,192	0.17	0.11	0.09	0.59	0.30	0.23	0.16	0.11	0.36	0.24	0.14	0.47	0.29	0.18
2003	2,648,152	0.09	0.04	0.03	0.52	0.22	0.17	0.10	0.07	0.28	0.18	0.09	0.37	0.21	0.11
2004	2,674,868	0.10	0.05	0.04	0.57	0.28	0.21	0.13	0.09	0.33	0.21	0.12	0.42	0.24	0.14
2005	2,850,511	0.10	0.05	0.04	0.59	0.29	0.21	0.14	0.09	0.35	0.23	0.12	0.45	0.26	0.15
2006	2,920,345	0.08	0.04	0.04	0.53	0.27	0.21	0.12	0.09	0.32	0.20	0.11	0.40	0.22	0.13
2007	2,808,928	0.09	0.05	0.04	0.54	0.29	0.22	0.15	0.11	0.34	0.22	0.13	0.41	0.25	0.15
2008	2,721,590	0.06	0.03	0.02	0.64	0.41	0.35	0.28	0.25	0.46	0.34	0.27	0.52	0.36	0.28
2009	2,661,767	0.14	0.09	0.06	1.37	1.07	0.99	0.90	0.84	1.15	1.00	0.87	1.21	1.04	0.91
2010	3,013,045	0.39	0.33	0.31	1.66	1.37	1.28	1.14	1.05	1.48	1.33	1.14	1.53	1.38	1.20
2011	3,174,896	0.05	0.03	0.02	1.43	1.24	1.19	1.11	0.95	1.31	1.21	0.98	1.35	1.23	1.00
2012	3,346,460	0.05	0.03	0.02	1.59	1.38	1.33	1.13	0.27	1.48	1.24	0.29	1.51	1.25	0.30
2013	3,605,316	0.04	0.03	0.02	1.60	1.39	1.33	0.39	0.26	1.47	0.48	0.28	1.50	0.50	0.29
2014	3,744,973	0.03	0.02	0.01	0.40	0.19	0.14	0.09	0.07	0.28	0.20	0.09	0.32	0.21	0.10
Median	2,808,928	0.09	0.05	0.04	0.59	0.31	0.23	0.15	0.11	0.36	0.24	0.15	0.47	0.29	0.18

Table 14: *The suites of criteria limits used in Table 13. Each suite of criteria limits are applied to the set of situations listed in the first three columns.*

Record Type	Unit Treated	AI rate type	Suite1	Suite2	Suite3	Suite4	Suite5	Suite6
Ag	Acres	Normal	mean7sd	mean10sd	mean12sd	mean7sd or fixed1	mean10sd or fixed2	mean12sd or fixed3
Ag	Acres	High, Medium, Adjuvant	mean5sd	mean5sd	mean7sd	mean5sd or fixed1	mean5sd or fixed2	mean7sd or fixed3
Ag	Cuft, Lbs, Misc	All	mean7sd	mean7sd	mean8sd	mean7sd or fixed1	mean7sd or fixed2	mean8sd or fixed3
Non-Ag	Acres	All	mean7sd	mean7sd	mean8sd	mean7sd or fixed1	mean7sd or fixed2	mean8sd or fixed3
Non-Ag	Cuft, Lbs, Misc	All	mean10sd	mean10sd	mean12sd	mean10sd or fixed1	mean10sd or fixed2	mean12sd or fixed3
Non-Ag	Application	All	mean5sd	mean5sd	mean7sd	mean5sd or fixed1	mean5sd or fixed2	mean7sd or fixed3

Table 15: *The recommended outlier criteria to use for records with rates of use and their values for different record types, units treated, and AI rate types.*

Record Type	Unit Treated	AI Rate Type	fixed2	Mean Limit
Ag	Acres	Normal Rate	100	mean10SD
Ag	Acres	Medium Rate	500	mean5SD
Ag	Acres	High Rate	2,000	mean5SD
Ag	Acres	Adjuvant	100	mean5SD
Ag	Cubic Feet	Normal Rate	10	mean7SD
Ag	Cubic Feet	Medium Rate	10	mean7SD
Ag	Cubic Feet	High Rate	10	mean7SD
Ag	Pounds	Normal Rate	50	mean7SD
Ag	Pounds	Medium Rate	50	mean7SD
Ag	Pounds	High Rate	50	mean7SD
Ag	Miscellaneous	Normal Rate	30	mean7SD
Ag	Miscellaneous	Medium Rate	200	mean7SD
Ag	Miscellaneous	High Rate	500	mean7SD
Non-Ag	Acres	Normal Rate	5,000	mean7SD
Non-Ag	Acres	Medium Rate	5,000	mean7SD
Non-Ag	Acres	High Rate	5,000	mean7SD
Non-Ag	Acres	Water Site	20,000	mean7SD
Non-Ag	Cubic Feet	Normal Rate	50	mean10SD
Non-Ag	Cubic Feet	Medium Rate	50	mean10SD
Non-Ag	Cubic Feet	High Rate	50	mean10SD
Non-Ag	Cubic Feet	Water Site	10,000	mean10SD
Non-Ag	Pounds	Normal Rate	20	mean10SD
Non-Ag	Pounds	Medium Rate	20	mean10SD
Non-Ag	Pounds	High Rate	20	mean10SD
Non-Ag	Pounds	Water Site	20	mean10SD
Non-Ag	Miscellaneous	Normal Rate	20,000	mean10SD
Non-Ag	Miscellaneous	Medium Rate	20,000	mean10SD
Non-Ag	Miscellaneous	High Rate	20,000	mean10SD
Non-Ag	Miscellaneous	Water Site	200,000	mean10SD

Table 16: *The recommended outlier criteria for pounds of AI per application for non-agricultural records and their values for different AI rate types and site types.*

AI Rate Type	Site Type	fixed2	Mean Limit
High Lbs AI 5	Other Site	100	mean5SD
High Lbs AI 4	Other Site	3,000	mean5SD
High Lbs AI 3	Other Site	10,000	mean5SD
High Lbs AI 2	Other Site	50,000	mean5SD
High Lbs AI 1	Other Site	200,000	mean5SD
Adjuvant	Other Site	5,000	mean5SD
High Lbs AI 5	Water Site	2,000	mean5SD
High Lbs AI 4	Water Site	10,000	mean5SD
High Lbs AI 3	Water Site	40,000	mean5SD
High Lbs AI 2	Water Site	60,000	mean5SD
High Lbs AI 1	Water Site	250,000	mean5SD
Adjuvant	Water Site	20,000	mean5SD

Bibliography

Barnett, V. and T. Lewis. *Outliers in Statistical Data*, 3rd edition. John Wiley and Sons, NY, 1994

Berk, Richard. *Statistical Learning from a Regression Perspective*. Springer, 2008.

Breiman, Leo. *Classification and Regression Trees*. Wadsworth, Pacific Grove, CA, 1984

Friedman, J. H. and Fisher, N. I. Bump hunting in high-dimensional data. *Statistics and Computing*, 9:123-143, 1999.

Hastie, T., Tibshirani, R. and Friedman, J. *The elements of statistical learning*, Springer, New York, 2009.

Hoaglin, D.C., F. Mosteller, and J.W. Tukey. *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, 1983.

Wilcox, R.R. *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*, Springer, 2010.

Wilhoit, L. *A Computer Program to Identify Outliers in the Pesticide Use Report Database*, Department of Pesticide Regulation, Report PM 98-01, 1998.