Department of Pesticide Regulation

Brian R. Leahy
Director

M E M O R A N D U M

Edmund G. Brown Jr.
Governor

TO: Edgar Vidrio
Senior Environmental Scientist (Supervisory)
Environmental Monitoring Branch

FROM: Colin R. Brown *Original Signed by*
Environmental Scientist
Environmental Monitoring Branch
916-445-9579

DATE: January 5, 2016

SUBJECT: EVALUATION OF 2014 USE DATA FOR 1,3-DICHLOROPROPENE SUBMITTED BY DOW AGROSCIENCES

1 Key Issue

In March 2015, Dow AgroSciences (DAS) submitted 2,886 records for 1,3-dichloropropene (1,3-D) use in California in 2014. On August 21, 2015, the Department of Pesticide Regulation's (DPR's) Pesticide Use Report (PUR) database yielded 2,758 records for 1,3-D use in 2014. In past years, the records submitted by DAS lacked the information necessary to create a one-to-one match between records across datasets, which resulted in a coarser comparison between the two datasets on the basis of Public Land Survey System (PLSS) township areas. The 2014 data submission by DAS includes additional data fields specifying an applicator permit number and a site application code for each record; these fields are also included in the PUR database and allow a more precise comparison of the two datasets.

Here, we develop methods to match individual records across PUR and DAS datasets. We then search for inconsistencies in PLSS section, county, and 1,3-D application amount between matched records. Additionally, we consider the special case of PLSS townships with a record of 1,3-D use that also overlap with two or more counties, as these townships may be more susceptible to accounting issues due to management responsibilities that are split among more than one county agricultural commissioner (CAC).

2 Background

1,3-dichloropropene is a fumigant used to control soil-borne pests in a variety of crops. In 2014, most 1,3-D use in California occurred in the San Joaquin Valley and Central Coast regions. In 1990, all permits for the use of 1,3-D were suspended in California. This action was based upon the results of limited monitoring studies in one high use county that indicated potentially high risk of cancer if some of the detected inhalation exposure levels persisted over the long-term (70 years). Following this action, DAS conducted several years of research to reduce exposure to handlers and bystanders, and proposed mitigation measures. Implementing the new use practices

to reduce ambient air exposure in combination with limits on the absolute amount of the fumigant used, DPR allowed the reintroduction of 1,3-D in 1995. The regulatory requirements on use practices and limitations on application amounts have been modified several times since 1995. Since 1999, the key mechanism that has been used to restrict use has been a cap on 1,3-D use within each township (6x6 mile area) of 90,250 adjusted total pounds per year, which is administered by DAS as a requirement for continued registration. However, use is allowed above the cap in townships where use since 1995 has been significantly under the amount allowed by the cap. This increase in annual use is limited to a total of 180,500 pounds, twice the 90,250-pound cap.

3 Evaluating and matching data from PUR and Dow AgroSciences

Our goal was to find a one-to-one match between records in each dataset on the basis of a primary key (unique identifier) or a composite key (multiple fields used together to identify a unique record), and to then analyze the paired data for discrepancies in PLSS township, county code, and reported application amount. Assuming a one-to-one relationship among records, the maximum match rate for the PUR dataset would be 100% (if all 2,758 records are positively matched), whereas the maximum match rate for the DAS dataset would be limited to 96% (2,758/2,886) as a result of the lesser number of records within the PUR dataset. We took an exploratory, iterative approach to matching records using algorithms that we developed based on our observations of the dataset. Our approach is summarized in Table 1 and detailed more extensively in the following paragraphs.

We began the matching process with the goal of creating a primary key for each record based on concatenation of a 7-digit applicator permit number with an 8-digit site-specific ID code. If a single record in each dataset shares this identical key, it forms the basis for a positive match. We filtered the data to remove any spaces, and padded permit and site numbers with leading zeros where necessary. We found more than 100 instances where a 15-digit key was shared by more than one record, and over 1,000 instances where inconsistencies in the permit number or site ID code between the DAS and PUR datasets prevented a match based on a 15-digit key. We appended an additional 4-character string to each 15-digit key to correct for the issue of duplicate primary keys. This 4-character string consisted of the numerical month of application (ranging from 01-12) and the first two digits of the application amount as listed in the DAS data. We used only the first two digits of the application amount, as we observed small discrepancies in the reported amount between datasets (these differences are further described later in this section). Once a 19-digit key was created for both the PUR and DAS datasets, we matched records based on exact matches in the ID key. Using this method we found positive matches for 1,497 records, or 52% of records in the DAS dataset.

We next designed filters to detect and account for systematic errors in the 19-character key. Here we define ’systematic’ errors as inconsistencies in the 19-character key that prevent an exact match. These errors may occur in either dataset and follow a predictable pattern (e.g. the transposition of the 2nd and 3rd characters in a string) and can be corrected through an automated process once identified. Once these errors are corrected, the resulting 19-character key can form the basis for a positive match.

We designed each filter based on our observations of patterns in the data. We found several systematic errors that could be corrected using simple formulas. For example, some 5-character strings in the site ID code field of the DAS database had been erroneously converted from a number to a date format. Accounting for this error, we gained an additional 65 positive matches. We found another systematic error in some records originating in Monterey County wherein the 15th character of the PUR record was transposed to the 11th character of the corresponding DAS record, present in 19 records. We verified accuracy in each match on the basis of identical application dates or application amounts.

Next, we took into account PLSS sections that were known to intersect county borders. Inconsistent reporting of county code by applicators can lead to mismatches in the dataset, because the first two digits of each permit code are defined by the county in which the application takes place. We hypothesized that errors resulting from inconsistent reporting of county would be more common in PLSS townships that intersect with one or more counties. We created a process by which unmatched records falling within a township containing two or more counties would be tested for positive matches based on a 19-character key using the alternative county codes for that township. This process yielded an additional 47 positive matches.

Table 1: Summary of the 9-step process used to match data between the PUR and DAS datasets.

Step	Filter Method	Method Description	Matches
1	19-character key	Permit No(7), Site No(8), Month(2), Amt(2)	1,497
2	Excel date code error	Numbers in format of 42xxx converted to a 3 digit date code	65
3	Correction for transposed characters	15th character of PUR key transposed to 11th character of DAS key	19
4	County code errors	Match based on alternative county codes for records in PLSS townships that intersect 1 or more counties	47
5	First 17 characters	Match based on similarities aside from application amount	629
6	Middle 13 characters	Broad match discarding county code, application date, application amount	67
7	Application amount	Match records according to nearest application value based on systematic differences of -0.62%, 0.87%, or 0% between the values reported in each dataset, followed by verification	256
8	Manual matches	According to similarities in date, application amount, character string	287
9	Validation Steps	Check for discrepancies in application amount (+/-1%) and application date (+/- 1day)	-142

Table 2: Comparison of records matched between the PUR and DAS datasets.

	DAS Data	PUR Data
Total Records	2,886	2,758
PLSS Townships in Data	328	313
Positive Matches	2,725	2,725
Match %	94.4%	98.8%
Unmatched Records	161	33

We next filtered the remaining unmatched records by searching for positive matches based on the first 17 characters of each 19-character string. In this way, we search for matches based on similarity in county, permit number, application month, but not necessarily a match in application amount. Such a process yielded an additional 629 possible matches. We then used a similar but broader approach to match records based on a 13 character string falling in the middle of each ID string, which disregarded information on county code, application month, or application amount. This final step yielded an additional 67 possible matches, for a total of 2,434 possible matches (84% of records in the DAS dataset).

We based our last series of filters on matches in reported application values. Based on observations of previously matched records, we found matched values most commonly differed (in descending order of frequency) by -0.62%, 0.87%, and 0.0%. These differences result from a slight difference in how DAS and PUR databases calculate the application amount (Neal 2014). We performed the filter in a three-step process by which the product of a multiplier and the DAS record application value would be tentatively matched to the record in the PUR dataset with the most similar application value. Each tentative match underwent a comparison of county code and search for similarities in the last 13 characters of each character string. If no positive match was identified, the process would repeat again for the next multiplier until a possible match had been found or until all three multipliers had been attempted. An additional 256 potential matches were identified in this manner (2,690 possible matches, or 93%).

We performed a manual search for any remaining unmatched records. We performed this step last due to the time consuming nature of the process. The manual search involved searching for records sharing short character strings, followed by a comparison of application date, application amount, and PLSS township. We matched an additional 287 records in this way.

We performed a validation process by comparing the application date and amount information for each tentative match. We flagged all matches wherein dates between records varied more than 1 day or application amount differed by 1%. We manually searched the PUR database for possible matches to each flagged DAS record, either finding probable matches based on application date and amount, or marking the record as having no match. Additionally, we manually searched for erroneous many-to-one matches and removed these from the final list. We deleted 142 erroneous matches as a result of this validation process.

4 Results

We matched 2,725 records between datasets using the 9-step filtering and validation process described above. The results of this process are summarized in Table 2. Those 2,725 records accounted for 98.8% of that data contained with the PUR dataset and 94.4% of that data contained within the DAS dataset. Thirty-three records remained unmatched in the PUR dataset, and 161 records remained unmatched within the DAS dataset. We found explanations for 43 of the 161 unmatched records in the DAS dataset: 40 records had not been received by the PUR from Del Norte County, and an additional 3 records contained negative values. Among matched records, 13 pairs contained a mismatch in the reported PLSS townships (Table 3). The location of each conflict is shown in Figure 1. Discrepancies in PLSS section resulted in the miscategorization of 47,162 adjusted lbs of 1,3-D statewide in 2014. The maximum value for any one misplaced record was 8,426 lbs, and the lowest value was 1,229 lbs. About half of the mismatched townships were adjacent to one another, whereas the remaining pairs were separated by one or more townships. Only 3 pairs were separated by a gap of two or more townships (Figure 1; pairs 11, 12, and 13).

Reporting errors in county code were more common than errors in the PLSS township, and errors in county code reporting did not necessarily lead to mistakes in township reporting. We identified 47 out of 2,886 records as having mistakes attributable to county code errors, and two of these records also showed discrepancies in the reported township. Our observations did not indicate that township mismatches were more likely to occur in townships that intersected county lines, although we did not verify this statistically. Of the 13 pairs containing mismatched townships, 8 pairs included at least one township that crossed county lines. The remaining 7 mismatches did not involve any townships along county lines. Figure 1 provides a view of the distribution of these mismatched pairs throughout the state.

54 townships exceeded the 90,250 adjusted lb cap for 1,3-D use in the DAS dataset, whereas 48 exceeded the 90,250 adjusted lb cap in the PUR dataset (Figure 2). The 6 additional townships exceeding the cap in the DAS dataset were the result of records not included in the PUR dataset, rather than misattribution of PLSS township associated with a given record.

Table 3: Summary of matched records containing mismatched PLSS township information, listed alongside county of record. We found 13 mismatched pairs in total, two of which also contained mismatched county information.

DAS Data		PUR Data	
Reported PLSS Township	Reported County	Reported PLSS Township	Reported County
16N 03E	Sutter	14N 03E	Sutter
15S 03E	Monterey	14S 02E	Monterey
03S 13E	Stanislaus	04S 13E	Stanislaus
14S 02E	Monterey	14S 04E	Monterey
05S 08W	Amador	05S 11W	Orange
15S 03E	Monterey	14S 02E	Monterey
15S 03E	Monterey	14S 02E	Monterey
09N 34W	Santa Barbara	10N 33W	Santa Barbara
10N 35W	Santa Barbara	11N 35W	San Luis Obispo
14S 23E	Fresno	15S 23E	Fresno
26S 10E	Tulare	21S 26E	Tulare
24S 29E	Tulare	16S 24E	Tulare
25S 24E	Tulare	19S 25E	Tulare

A single township exceeded the 180,500 adjusted lb usage cap (Figure 2), but not as a result of a township mismatch. The township exceeded the cap in the PUR record, by approximately 800 adjusted lbs, but was under the cap by approximately 300 adjusted lbs in the DAS record. The excess was attributable to the accumulation of small errors across several records rather than discrepancy in township or missing records. The DAS and PUR use different methods of calculating active ingredient (AI) from the product composition, resulting in small errors (usually less than 1%). These inconsistencies have been previously described by Neal (2014).

Reported use for 1,3-D statewide was higher overall in the DAS as compared to the PUR, and was generally higher on a township basis (Figure 3). Statewide use for 1,3-D totaled 13.8 million lbs (14.6 million lbs adjusted) according to the DAS, whereas the PUR database reports a total of 13.1 million lbs (14.0 million adjusted lbs), about 5% less than the DAS. Additionally, the DAS dataset contained 14 additional PLSS townships not found in the PUR database, contributing to the large deficits in PUR 1,3-D accounting observed in Figure 3. Most of these differences are attributable to the 161 records present in the DAS dataset that are missing from the PUR, rather than due to discrepancies between the reported usage contained within the 2,725 instances of paired data; the DAS agrees with the PUR within 0.5% of usage statewide when only considering paired data.

We did find substantial differences between our list of multi-county PLSS townships and the list provided by DAS. We identified 77 PLSS townships with record of 1,3-D usage that fell into two or more counties, whereas the list provided by DAS identified only 33. The DAS list contained 2

townships that did not appear on our list; however, these townships were not multi-county townships by our analysis. A single additional township appeared on the DAS list of multi-county townships with no 1,3-D use in 2014, but this township was associated with 1,3-D use in the larger PUR and DAS datasets that we used in the matching process. Overall, 31 townships appeared on both lists and our list contained 45 multi-county townships that did not appear on the DAS list. Our list of multi-county townships is provided in Table 4.

Based on our evaluation of the submitted data, we recommend that DAS take several steps to facilitate the comparison of the DAS and PUR datasets:

- DAS could provide alphanumeric grower ID or numeric crop commodity codes. These fields are already found in the PUR and would provide an additional means of creating a unique key for each record.
- DAS could change or make transparent the method by which application amounts are calculated. Ideally, DAS could use the same method used by PUR in calculating application amount. In the current scenario, reported values from DAS can be lesser than, equal to, or greater than those values reported by PUR. Such inconsistencies introduce a greater degree of complexity into the matching process.
- DAS could fix certain transcription errors, such as conversion of some 5-character numeric strings in the permit ID field to a 3- or 4-character date code.

Implementation of the above recommendations should help improve the accuracy and speed of the matching process in future years.

cc: Dr. Marylou Verder-Carlos, DPR Assistant Director
Mr. George Farnsworth, DPR Assistant Director
Mr. David Duncan, DPR Environmental Program Manager II
Mr. Randy Segawa, DPR Special Advisor, Pesticide Programs Division
Ms. Pam Wofford, DPR Environmental Program Manager I

References

Neal, R. (2014). Evaluation of "2013 pesticide use data of adjusted pounds 1,3-dichloropropene in California", submitted by DOW Agrosciences on 4/25/14. Sacramento, CA: Department of Pesticide Regulation, California Environmental Protection Agency.

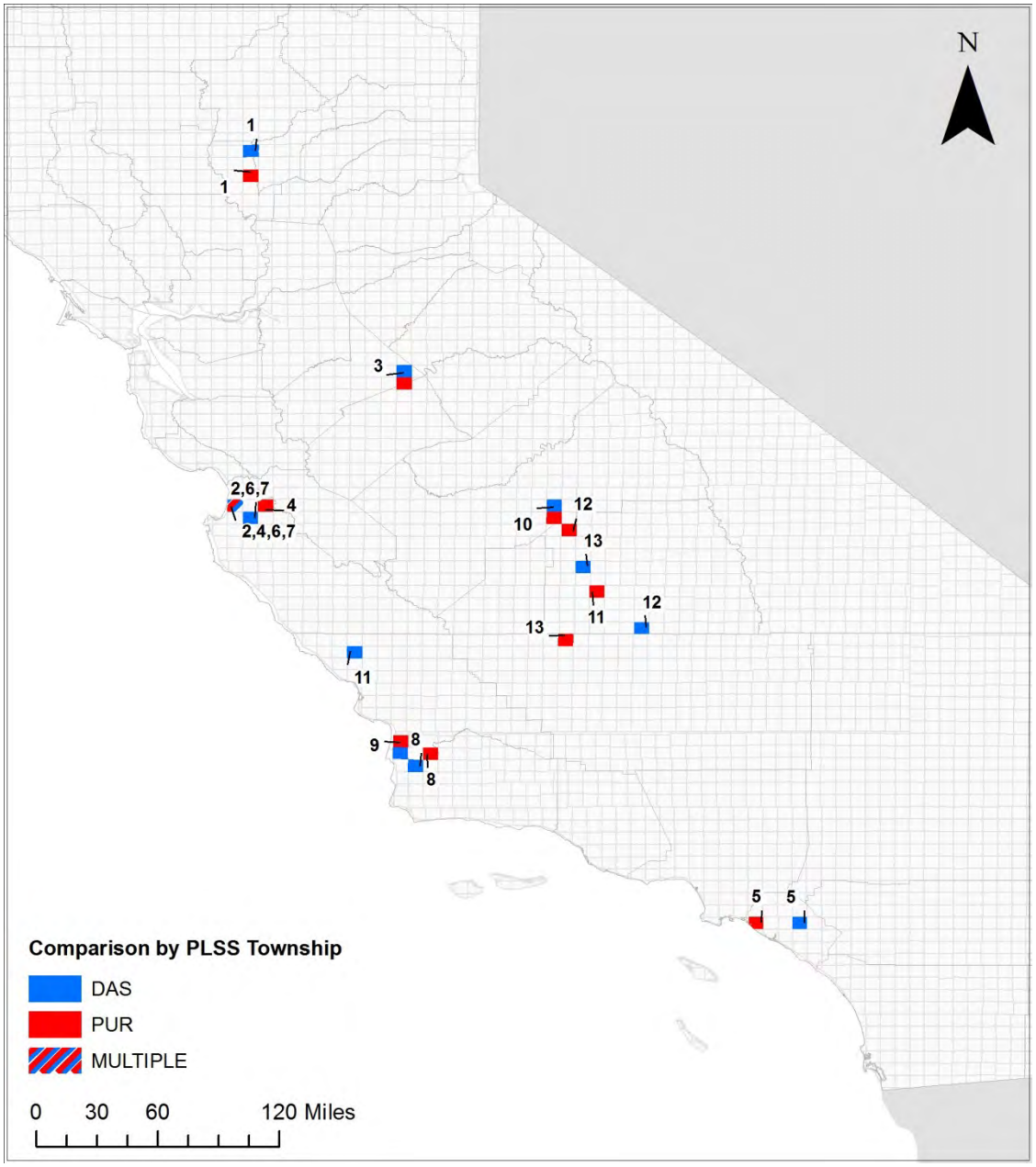

Figure 1: Mismatches presented by pair from comparison of the 2014 DAS and PUR datasets. We found a total of 13 mismatched pairs. Only in 2 cases did pairs not correspond in county attribution.

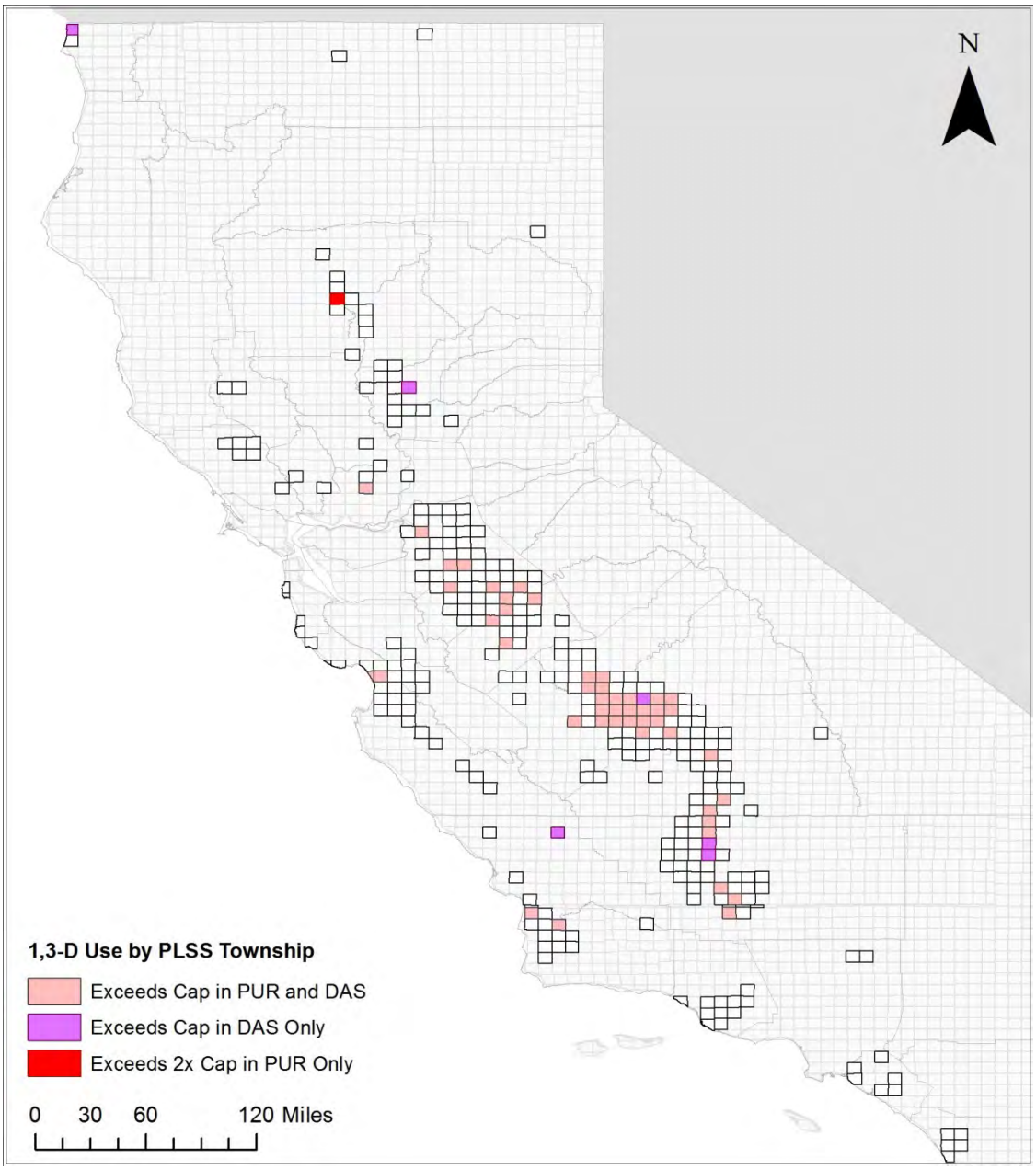

Figure 2: Map indicating PLSS townships that exceeded the cap for 1,3-D usage in 2014. 54 townships exceeded the 90,250 adjusted lb cap, 6 of which only occurred in the DAS dataset. A single township exceeded the 2x (180,500 adjusted lb) cap in the PUR dataset, but not the DAS.

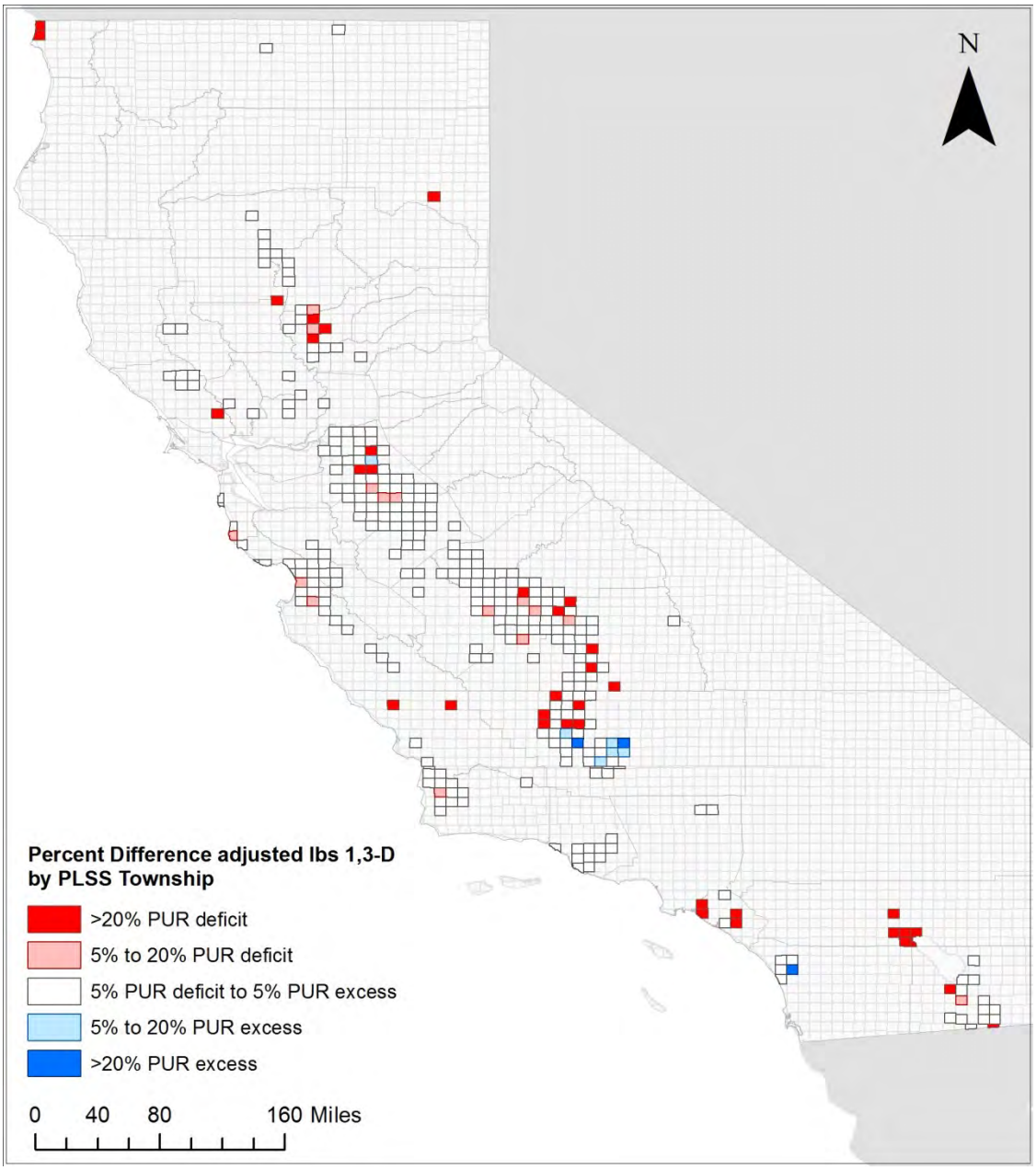

Figure 3: Comparison of 1,3-D use by township presented in terms of how the PUR differs from the DAS by township. We found that the higher total use reported in the DAS dataset was attributable mostly to the greater number of records in the DAS dataset, as opposed to discrepancies among records present in both datasets.

Table 4: List of multi-county townships with a record of 1,3-D use in 2014. We created the list using an overlay of county and township shapefiles in ArcGIS, and selecting for those townships with 1,3-D use in 2014 that also intersected two or more county lines. We found 77 townships that met this criteria, and the description of each is provided below.

Township	Count	County 1	County 2	County 3	County 4	1,3-D adj. lbs (PUR)	1,3-D adj. lbs (DAS)
M23N02W	3	BUTTE	GLENN	TEHAMA		181,327	180,210
M13S17E	2	FRESNO	MADERA			176,727	175,639
M12S02E	2	MONTEREY	SANTA CRUZ			170,678	173,004
S11N35W	2	SAN LUIS OBISPO	SANTA BARBARA			167,820	167,774
M04S12E	2	MERCED	STANISLAUS			151,446	150,513
M13S18E	2	FRESNO	MADERA			149,007	148,089
M15S23E	2	FRESNO	TULARE			142,101	146,148
S10N33W	2	SAN LUIS OBISPO	SANTA BARBARA			134,782	134,568
M16S22E	3	FRESNO	KINGS	TULARE		129,387	159,612
M12S18E	2	FRESNO	MADERA			122,595	121,840
M02S08E	2	SAN JOAQUIN	STANISLAUS			114,299	114,185
M12S01E	2	MONTEREY	SANTA CRUZ			112,394	112,176
M17S21E	2	FRESNO	KINGS			99,578	98,964
M02S07E	2	SAN JOAQUIN	STANISLAUS			96,867	96,470
M05S11E	2	MERCED	STANISLAUS			93,481	92,906
S10N35W	2	SAN LUIS OBISPO	SANTA BARBARA			89,486	89,619
S10N34W	2	SAN LUIS OBISPO	SANTA BARBARA			88,505	88,662
M05S10E	2	MERCED	STANISLAUS			83,418	86,868
M21S18E	2	FRESNO	KINGS			81,638	83,837
M06S10E	2	MERCED	STANISLAUS			80,905	80,406
M17S20E	2	FRESNO	KINGS			75,221	74,757
M13S16E	2	FRESNO	MADERA			74,027	73,571
M17S22E	3	FRESNO	KINGS	TULARE		70,013	69,581
M12S12E	2	FRESNO	MERCED			68,245	67,825
M16S23E	2	FRESNO	TULARE			66,545	67,108
S10N26W	2	SAN LUIS OBISPO	SANTA BARBARA			62,919	62,532

Continued on next page

Table 4 – continued from previous page

Township	Count	County 1	County 2	County 3	County 4	1,3-D adj. lbs (PUR)	1,3-D adj. lbs (DAS)
M17N03E	3	BUTTE	SUTTER	YUBA		60,935	64,177
S11N34W	2	SAN LUIS OBISPO	SANTA BARBARA			58,163	58,178
M09S15E	2	MADERA	MERCED			50,890	50,576
M05S12E	2	MERCED	STANISLAUS			49,968	49,985
M15N03E	2	SUTTER	YUBA			45,944	54,485
M23N01W	3	BUTTE	GLENN	TEHAMA		42,741	42,478
M11S02E	2	SANTA CLARA	SANTA CRUZ			41,470	41,785
M16N03E	2	SUTTER	YUBA			38,926	57,991
M02S10E	2	SAN JOAQUIN	STANISLAUS			38,084	37,850
M04N06E	2	SACRAMENTO	SAN JOAQUIN			37,725	37,492
M06S09E	2	MERCED	STANISLAUS			32,390	32,190
M11S04E	2	SAN BENITO	SANTA CLARA			30,951	32,248
M02S09E	2	SAN JOAQUIN	STANISLAUS			29,899	29,852
M12S11E	2	FRESNO	MERCED			28,220	28,046
M13S03E	2	MONTEREY	SAN BENITO			28,185	28,964
M15S24E	2	FRESNO	TULARE			27,453	27,284
M12S03E	4	MONTEREY	SAN BENITO	SANTA CLARA	SANTA CRUZ	27,258	27,436
M18S23E	2	KINGS	TULARE			25,904	25,758
M14N03E	2	SUTTER	YUBA			25,777	29,342
M14S04E	2	MONTEREY	SAN BENITO			23,658	23,817
M13N03E	2	SUTTER	YUBA			22,573	22,467
M03S07E	2	SAN JOAQUIN	STANISLAUS			22,233	22,096
M02N04E	2	CONTRA COSTA	SAN JOAQUIN			22,195	22,058
M12S19E	2	FRESNO	MADERA			20,577	20,450
M13S19E	2	FRESNO	MADERA			19,946	19,823
M03S13E	2	STANISLAUS	TUOLUMNE			18,952	21,573
M13N05E	3	PLACER	SUTTER	YUBA		17,362	17,255
M17S34E	2	INYO	TULARE			16,547	16,445
M03S06E	2	SAN JOAQUIN	STANISLAUS			15,694	15,597
M17N02E	2	BUTTE	SUTTER			15,372	15,277

Continued on next page

Table 4 – continued from previous page							
Township	**Count**	**County 1**	**County 2**	**County 3**	**County 4**	**1,3-D adj. lbs (PUR)**	**1,3-D adj. lbs (DAS)**
M12S20E	2	FRESNO	MADERA			13,357	13,274
M12S04E	3	SAN BENITO	SANTA CLARA	SANTA CRUZ		12,898	12,837
M22N02W	3	BUTTE	GLENN	TEHAMA		11,453	11,383
M04S13E	2	MERCED	STANISLAUS			10,755	10,689
M13N04E	3	PLACER	SUTTER	YUBA		9,420	9,362
M07N04E	2	SACRAMENTO	YOLO			9,076	9,020
M11S03E	2	SANTA CLARA	SANTA CRUZ			7,041	7,103
M07S08E	2	MERCED	STANISLAUS			6,678	6,637
M09S04W	2	SAN MATEO	SANTA CRUZ			6,416	6,376
M12N03E	2	SUTTER	YUBA			6,165	6,127
M21S17E	2	FRESNO	KINGS			4,738	4,663
S04S11W	2	LOS ANGELES	ORANGE			4,446	5,447
M12S14E	2	FRESNO	MADERA			4,174	4,148
M04S06E	2	SAN JOAQUIN	STANISLAUS			4,013	3,988
M07S09E	2	MERCED	STANISLAUS			3,210	3,191
M24N02W	2	BUTTE	TEHAMA			2,387	2,373
M04N05E	2	SACRAMENTO	SAN JOAQUIN			2,337	2,323
M15N01E	2	COLUSA	SUTTER			2,003	1,990
M10N08W	2	LAKE	SONOMA			1,402	1,393
S03S09W	3	LOS ANGELES	ORANGE	SAN BERNARDINO		671	671
M08N02E	2	SOLANO	YOLO			427	427