Final Report on the Accuracy and Feasibility of Using Spatial Data to Refine the California Pesticide Use Reporting Data Set

Submitted to Impact Assessment, Inc., Berkeley, CA

May 28, 1999

Stephanie J. Weigel, John R. Nuckols, Chris C. Skinner, and Ryan S. Miller

Environmental Health Advanced Systems Laboratory
Department of Environmental Health
Colorado State University
Fort Collins, CO

I. ACKNOWLEDGEMENTS

Funding for this project was provided by a contract with Impact Assessment, Inc. acting on behalf of the California Department of Health Services, Environmental Epidemiology and Geographic Information Section, Oakland, California, Peggy Reynolds, PhD, Director, as part of NCI grant #CA717450. Assistance in the production and preparation of the report was provided by Dr. Reynolds and Martha Harnly of the CDHS, and by Robert Gunier and Andrew Hertz of Impact Assessment, Inc. Assistance of the California Department of Water Resources is also acknowledged.

Disclaimer:
The contents of this report was prepared by a research institution independent of both the California Department of Health Services and Impact Assessment, Inc. The findings and opinions expressed by this report may not necessarily represent the viewpoints of CDHS or Impact Assessment, Inc.

TABLE OF CONTENTS

IV. LIST OF TABLES AND FIGURES

III. EXECUTIVE SUMMARY

Pesticide-related risks have become a topic of increasingly important public health concern in the state of California. Due to the large amount of agriculture that occurs in the state, many residences are located in or near agricultural areas that are subject to intensive agricultural use, and residents may be exposed to agricultural chemicals by proximity to these intensively cultivated areas. To better refine exposure assessments for populations potentially exposed to agricultural chemicals, a pilot project was undertaken by the Environmental Health Advanced Systems Laboratory in collaboration with the California Department of Health Services, with the objectives of: 1) validating and evaluating the California Pesticide Use Reporting (PUR) database and available spatial data sets representing locations of agricultural activities for use in refining exposure assessments; and 2) assessing the use of data sources from Objective 1 to determine their utility in future epidemiological studies performed over large spatial and temporal extents. Procedures for identification of formerly agricultural lands were also assessed as part of the second objective.

The PUR database was validated by comparing acres of pesticide application reported in PUR with a spatial database developed by the California Department of Water Resources (CDWR). The CDWR database is a spatial database containing 83 land cover classifications at a resolution of 2 acres. The data was derived from aerial photography, and is100% ground truthed by field crews. The purpose of the database was to study patterns in water consumption. Twenty-three (23) different surveys have been performed in twenty (20) counties, mostly in the California Central Valley. The PUR and the CDWR datasets showed similarities in reported acres for different crops, however some discrepancies were noted, possibly due to seasonal factors, multiple cropping, or lack of pesticide application to certain crops. For Kings County, cotton was the highest ranked crop in terms of number of acres for both CDWR and PUR, as well as the highest rank in pounds applied. The PUR database represented 99% of CDWR acreage. The next four highest acreage crops in the PUR database, alfalfa, safflower, cotton and tomatoes, were 72%, 84%, 39% and 124% of CDWR acreages, respectively.

The California GAP analysis project database was examined for use in refining agricultural pesticide use locations beyond the level that is currently available in the PUR, this is a section in the Public Land Survey System (PLSS, approximately 1.0 mi^2). GAP data are part of a nationwide program to assess habitat suitability for different wildlife species (Scott and Jennings 1997).

Analysis for California is complete, and was conducted by the United States Fish and Wildlife Service (USFWS) and the University of California, Santa Barbara. The GAP land use classification system is based on a modified Holland and Anderson classification scheme (Anderson et al. 1976, Holland 1986; Davis, 1998). GAP classifies 42 general land-use/land-cover types for the purpose identifying critical habitat types. The primary spatial data source for the GAP was Landsat Thematic Mapper data from the early 1990s obtained at a pixel resolution of approximately 30 m^2. For the purposes of our study, CDWR and GAP were mapped to a common classification scheme. The CDWR data set was reclassified to simulate the classification system of agricultural cover types used by GAP. The resultant classification scheme identified 12 basic land use practices, with an emphasis on agricultural land. The results of our study indicate GAP to be reasonably accurate in locating coarse agricultural classes of land use for the two counties and time periods studied (Kings 1991 and San Joaquin 1988). The results show that at this level of classification, GAP overestimated the total acres in the class agricultural land in both counties (San Joaquin ~10% and Kings ~21%) and the class riparian vegetation in San Joaquin county (~306%). GAP underestimated the total acres in all other cover classes. We examined the effect of aggregating land cover to seven (7) categories, including only two major subcategories related to agricultural production: general agricultural (row and field crops, grain crops, mixed barren land, rice, and pasture) and orchard/vineyards. At this level of resolution, classification accuracy for agricultural land was good, achieving approximately 81% identification accuracy for both counties. Classsifcation accuracey for orchards/vineyards was fair - 78% for San Joaquin county and 72% for Kings county.

Procedures were analyzed for increasing the spatial resolution of the PUR data set using ancillary spatial data such as GAP or CDWR. At this time, GAP is the only statewide data set available, however other statewide data sets may be available in the near future. This preliminary work demonstrates the technical feasibility of using a higher spatial resolution data set to enhance PUR section level data. Preliminary analysis shows that about 34% of the township and range sections in Kings County for 1991 would see an improvement in the resolution of exposure classification where a spatial merge with GAP is utilized. Most of these sections are located near urban and rural regions, which is where much of the population is located in Kings County. In cases where the majority of cases and controls reside in these areas, using GAP in the spatial merge would be of benefit to an epidemiological investigation. Using a higher resolution data set such as the CDWR data would allow for increased exposure resolution in 76% of the counties in Kings County. Further study could elucidate more precisely the types of land uses that would be spatially refined, as well as their typical location in the landscape (e.g. peri-urban areas). Since a complete county-level overlay would be quite analytically extensive, even studying a subset of sections of a county would greatly assist in the development of a fully developed exposure assignment protocol, by demonstrating the utility of higher spatial resolution exposure information. This process could be assisted in the future by the availability of new statewide land use/land cover data sets.

Different methods for refining exposure information using Geographic Information Systems analytical and modeling techniques were presented as possibilities for future work. Several data sets were used to identify lands that are currently classed as urban, but were classified as agricultural in the mid-1970's. Overlays were performed using USGS Land Use Land Cover (LULC) data to represent mid-1970's agricultural areas with urban designations from both the California Department of Water Resources (CDWR, 1988 and 1991) and the California GAP analysis project (early 1990's) data sets for San Joaquin and Kings. Areas were thus delineated that are urban in the CDWR or GAP time period, but agricultural in the 1970s and when the LULC data were collected. The utility of point-in-polygon overlay for determining residence on former agricultural land was demonstrated, using LULC agricultural coverages and randomly generated point data to represent pseudo-subject point locations. The GIS could be used to generate input data for statistical analyses on these points, such as correlations of disease outcomes with types of crops in the vicinity. All of these analytical GIS techniques may assist in the evaluation of pesticide exposure histories for certain types of health outcomes.

In summary, the study investigated two major issues concerning refinement of pesticide exposures in California. We examined the Pesticide Use Reporting (PUR) database and available spatial data sets that may be used with it to refine exposure to pesticides. We established a procedure that would allow the identification of lands that were formerly agricultural. To validate the PUR database we compared the acres of pesticide application to a spatial database developed by the California Department of Water Resources (CDWR). Analyzing Kings County for 1991, we found similarities in reported acres of pesticide application per crop to the acres in the CDWR data, but certain discrepancies between CDWR and PUR acreages warrant future research. In particular, examining the data sets at the section level may prove useful in detailing many of the differences between CDWR and PUR. Additionally, the validation of PUR in other counties for other years would also shed light as to the accuracy of the data.

We validated the GAP data set (using CDWR) to determine its utility for merging with the PUR to refine the location of pesticide exposure. We found that the GAP was reasonably accurate in locating coarse agricultural classes of land use such as 'Agricultural Lands' and 'Orchard/Vineyard' for the two counties and time periods we studied (Kings, 1991 and San Joaquin, 1988).

We analyzed procedures for increasing the spatial resolution of the PUR data set. Both the GAP and the CDWR data sets can be used in conjunction with the PUR for this purpose, but only the GAP is appropriate where wide-scale application is needed, due to availability. In future work, the technique should be applied to an entire county (or a subset thereof) using subject residence point locations to ascertain the utility of using each of the data sets with PUR.

Another issue involved in using refined exposure information is determining the actual exposure to households. Point-in-polygon, buffer and exposure models all can be used for this, but each has a host of its own problems associated with it. Future work may also include the evaluation of these techniques in a small area in conjunction with household dust samples for validation.

V. INTRODUCTION

In the larger context of refining exposure assessments for populations potentially exposed to agricultural chemicals, a pilot project was undertaken by the Environmental Health Advanced Systems Laboratory (EHASL) in collaboration with the California Department of Health Services (CDOH). The objectives of this pilot project were as follows: 1) To validate the California Pesticide Use Reporting (PUR) database for use in refining exposure assessments for populations exposed to pesticides in California agricultural landscapes. This objective also involved the examination of spatial databases and procedures that may be used in conjunction with the PUR to refine pesticide exposures. 2) To assess the use of data sources from Objective 1 to determine their utility in future epidemiological studies that may be performed over large spatial (e.g. statewide) and temporal (e.g. ascertaining residence on former agricultural land) extents.

V. BACKGROUND

Pesticide-related risks have become a topic of increasingly important public health concern in recent years, particularly in the state of California. Many residences in California are located in agricultural areas that are subject to intensive pesticide use. Residents living in agricultural areas may be exposed to pesticides due to the proximity of their residences to treated cropland (Simcox et al., 1995; Richter, 1992). Contamination of soil, airborne particulate matter, and water supplies in agricultural areas results from normal pesticide applications, pesticide drift and over-spray (Goolsby et al., 1997; Maas et al., 1995; Camann, 1994). Questionnaires are not useful for ascertaining information about pesticides used near a residence unless the land was farmed by the respondent. Agricultural pesticides have been measured in carpet dust samples in homes at distances greater than one-quarter to one-half mile from cultivated fields (Bradman et al., 1997; Simcox et al. 1995). These exposures could not be explained solely by occupational "take-home" exposures, and levels were inversely correlated with the estimated distance of the residence from crop fields (Simcox et al., 1995).

Remote sensing and geographic information systems (GIS) have been used to study associations between landscape characteristics and the incidence of disease (Glass et al.1995, Beck et al. 1994). Satellite image data have been used to classify agricultural land by crop type (Ward and Nuckols 1999, Maxwell et al. 1996, Campbell, 1996). Land cover types (e.g. vegetation, bare soil, water, urban areas) differ in their reflectance and spectral reflectance characteristics, which allows the classification of satellite imagery into land cover types and individual crop species. The use of ancillary data, such as ground truthed cropping information (Ward and Nuckols, 1999), can improve the accuracy of the classification. A study by Nuckols et al. (1996) used this approach to investigate the importance of the proximity of maternal residences to specific crops as a risk factor for low birth weight. Because pesticide use varies by cultivation practices (Johnson and Kamble, 1984), crop type may provide a useful surrogate for possible exposures to pesticides applied to crops.

There are a number of available spatial data sets that classify crops by type. This work investigated several of those data sets, to ascertain their utility as exposure assessment data layers. Some of the data sets (e.g. PUR, CDWR) are available at a state level. Others, including GAP, LULC and MRLC, are discussed below and are available as part of national mapping programs. The feasibility of doing studies at the statewide level may be closely tied to the ability to obtain existing and appropriate spatial data showing the locations of agricultural land use.

This study is significant in that it develops methodologies for the assessment of population exposures by refining current methods of determining exposure. This is done by bringing together parallel efforts in the use of GIS tools in exposure assessment with health outcome research. In particular, this preliminary work provides a basis for evaluating the utility of using existing information on agricultural pesticide use in future studies of rare health outcomes in California.

VI. OBJECTIVE 1: VALIDATION AND ASSESSMENT OF THE PESTICIDE USE REPORTING (PUR) DATABASE

Introduction

The California Pesticide Use Reporting (PUR) database records information on regulated pesticides in the state of California including application rate, date of application, crop applied and the township-range section in which the application occurs. Before 1990, the reporting of all pesticide applications was not mandatory -- only a short list of restricted pesticides was recorded in the PUR. In 1990 it became a state law to report all regulated pesticide applications, with counties bearing the responsibility to compile the information. Additionally, 1990 and onward PUR databases implemented an expanded classification scheme and also included a unique field-identifier attribute. Unfortunately, PUR collected for 1990 was plagued with many problems due to the changeover in regulations, and is considered unusable for the purposes of this study.

The purpose of validating the PUR was to determine how closely it represents pesticide usage in the county and its utility in exposure assessment studies.

For the validation of the PUR, we used a land cover data set created by the California Department of Water Resources (hereafter referred to as the CDWR data set). Derived from aerial photography, the CDWR data is 100% ground truthed by field crews, and has 83 land cover classes defined. The department collects the data to study patterns in water consumption. Approximately six county-level land use surveys are conducted each summer. Thus far, 23 different surveys have been performed in 20 counties, mostly in the California Central Valley. The location of these surveys are listed in Table 1 and shown in Figure 1. The CDWR plans to conduct such a survey every 7 years for each county to develop an extensive temporal record of water consumption.

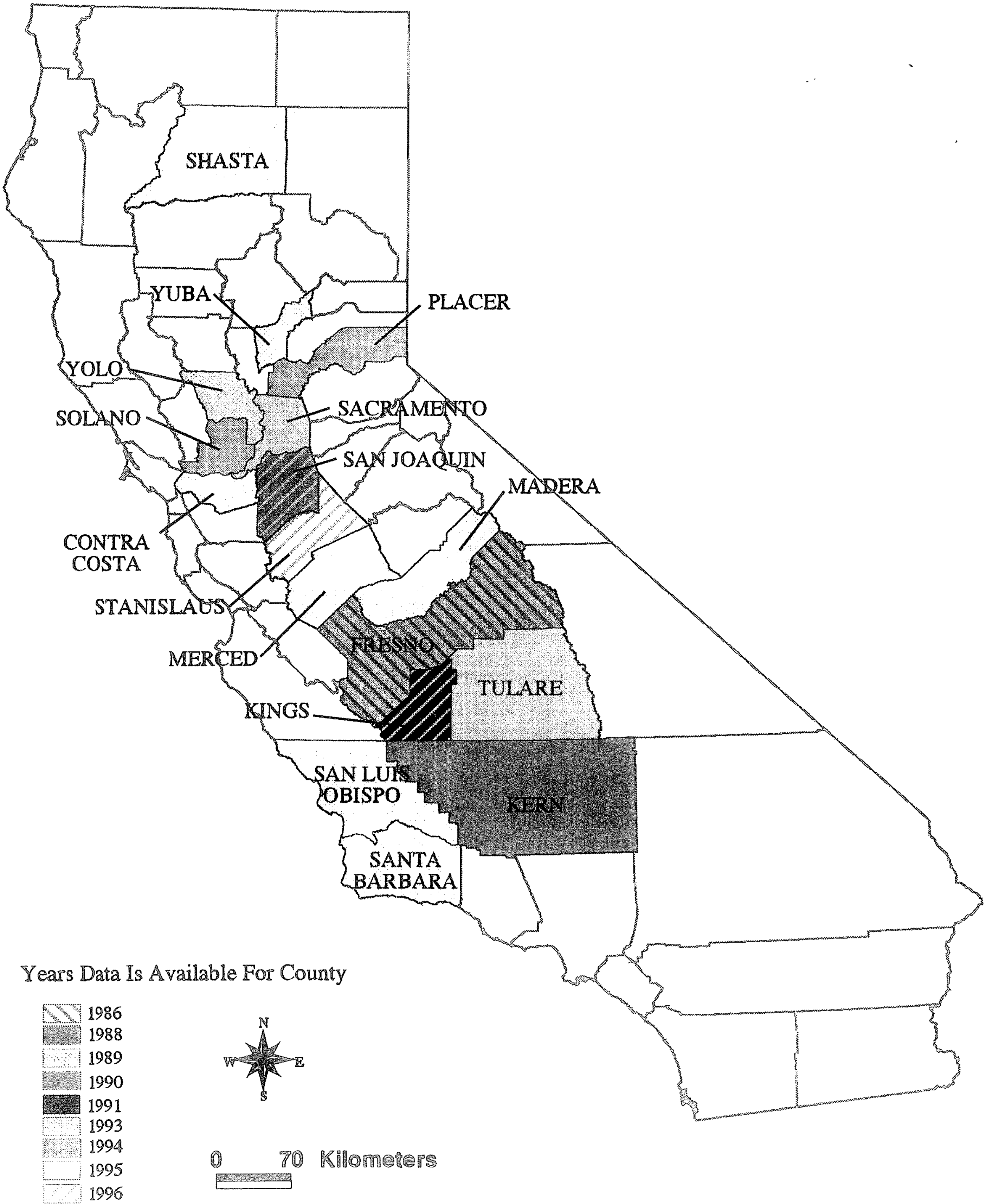
TABLE 1. AVAILABILITY OF CALIFORNIA DEPARTMENT OF WATER RESOURCES DATA.

Year for which data is available	Counties or other area
1976	Legal, Delta
1986	Fresno
1988	San Joaquin
1989	Yolo
1990	Kern
1991	Legal, Kings
1993	Sacramento, Tulare, and Upper Santa Ana river drainage area
1994	Placer, Solano, Fresno
1995	Contra Costa, Yuba, Shasta, San Luis Obispo, Santa Barbara, Madera, Merced
1996	San Joaquin, Kings, Stanislaus

Note: Data is available on CD (at a cost of $50.00 per CD) by writing Tom Hopkins at the CDWR at 1416 9th Street, Sacramento, CA 95834.

The format of the CDWR data can be cumbersome. We received data on CD for Kings (1991 and 1996) and San Joaquin County (1988 and 1996). The data was in a standard Arc/INFO (ESRI, Redlands, CA) export format (*.e00) for all but one of the data sets. The Kings County 1991 data was in an AutoCAD (Autodesk, Inc., San Rafael, CA) format, which required the use of a series of Arc/INFO Arc Macro Language (AML) scripts for conversion to Arc/INFO format.

Figure 1.
Available CDWR data for California counties



These AML's were provided on the CD, but the instructions provided were very complicated and required extensive error-checking at each step of the process.

Methods

The validation of the PUR consisted of comparing acres of pesticides applied to specific crops to the reported acreage of the crop in the CDWR database. To accomplish the comparison we had to first interpret the PUR to account for multiple pesticide applications to the same field. A C++ program was written to perform this task, which linked the pesticide applications using the field-identifier code. Since the acreage of each pesticide application could vary for the same field, we assumed the maximum acreage reported in the PUR was the maximum area of the actual field. For Kings County in 1991, this processing step reduced 51,980 pesticide applications to 3,716 separate fields. We could not use the San Joaquin 1988 data for this analysis because it lacked the needed field-identifier. From these results, we were able to generate descriptive statistics of the PUR dataset.

The next step was to match the 58 pesticide use codes (designating by crop type or land cover type) in the PUR with 83 CDWR codes (designating land use or land cover) which were found in Kings county in 1991. This step proved simple for crop classes, but more difficult where ambiguous land uses or pesticide applications were specified in either data base (e.g. 'Cropped within the last 3 years', 'Structural Pesticide Control', etc.). Appendix A contains a complete listing of the reclassification scheme we utilized.

Descriptive statistics of the PUR database were developed, and total acreage of pesticide use reported was compared for each crop relative to the acreage for each crop in the CDWR data set. Total acres of reported pesticide application by crop was also aggregated for PUR by Public Lands Survey Section (PLSS) section to examine the consistency of the reporting as compared with the standard 640 acre area of a PLSS section.

Results

After adjusting for multiple applications to fields, out of the 899 pesticide-applied sections in Kings County in 1991, 86 sections (~10% of the pesticide-applied sections) still reported more than 640 acres in a section. The standard acreage of most PLSS sections is 640 acres. Seventeen

(~2%) had more than 1000 acres of pesticide application reported. The maximum acreage reported of pesticide application for a single section was reported at 1796 acres.

Twenty-nine crop types were recorded in both the CDWR and PUR. Sixteen land cover types were found in the CDWR, but were not reported as having any pesticide use according to the PUR. Ten crops had some pesticide use reported in PUR, but were not found within the CDWR data set, including wheat, oats and broccoli. These crops are listed in Table 2, along with the rank of each record for CDWR acres, and PUR acres and total pounds of pesticide applied. Of note, barley, cherries and green beans have significantly higher acreage of pesticide applications reported than the acres delimited by the CDWR spatial data set.

Discussion

Table 2 ranks crops in terms of CDWR acres, PUR acres, PUR pounds of pesticide applied, and pounds of pesticide applied per acre, using CDWR acres. Crops that rank in the top ten for acreage (CDWR or PUR) or total pounds of pesticide applied are bolded, and crops that rank in the top ten pounds per acre are italicized. Fruit and nut trees and vineyards predominate in the highest ranked crops in terms of pesticide applied per acre[1], while cotton, alfalfa, sugar beets, tomatoes are some of the top crops in terms of acres and pounds applied.

There are number of reasons that explain the differences between the acreage reported in PUR and CDWR. The first is the time during the year that crops are grown. This is important in instances where the crop a pesticide was applied to may not have been present on the ground during the time frame in which the CDWR data was collected (usually summer). Thus, while the PUR is collected throughout a given year, the CDWR data provides a 'snap-shot' of crops on the landscape. Some crops, such as alfalfa, are harvested periodically and may not be captured by a one time 'snap-shot.' Some crops may be grown year round on a multiple cropping basis, and thus one field (as indicated by CDWR) may have several pesticide applications. By using only PUR applications from the summer, to more accurately coincide with the CDWR data, some of the suspected temporal mismatching with the CDWR data set may be reduced.

---

[1] Note that the CDWR acreages were used to calculate these values for pounds per acre, and thus the accuracy of the values depends on the accuracy of the CDWR acreages. We would expect CDWR acreages to vary most from actual values in cases where crops are seasonal or multi-cropped, since acreages were obtained by field estimators at one time point in the year.

Table 2: Comparison of acres cropped vs. acres applied with pesticide, Kings County 1991

Note: Top ten crops for Area and/or Lbs Pesticide Applied are bolded, top ten crops for Lbs/Acre are italicized

Description	CDWR Area, Acres	PUR Tot Lbs Pesticide	PUR Acres	%CDWR	Lbs PUR Pesticide Applied/Acre (CDWR)	Rank, CDWR Acres	Rank, Lbs Applied	Rank, PUR Acres	RANK, Lbs PUR/Acre (CDWR)
BARLEY	31.8	4394.9	4,955.2	15587.2%	138.25	42	21	9	2
CHERRIES	12.7	1910.2	75.6	594.8%	150.28	44	25	32	1
Green Beans	345.2	837.2	1,144.0	331.4%	2.43	31	31	17	25
APRICOTS	125.5	15500.5	276.5	220.4%	123.55	36	14	27	3
Lettuce	202.7	1145.2	320.0	157.9%	5.65	35	30	24	18
TOMATOES	11,075.6	348,421.2	13,764.3	124.3%	31.46	6	4	5	8
SUGAR_BEETS	5,731.4	603916.2	6,147.5	107.3%	105.37	10	2	8	4
Kiwis	281.6	1209.7	302.0	107.2%	4.30	32	29	25	20
COTTON	232,441.5	1708767.1	230,143.2	99.0%	7.35	1	1	1	17
APPLES	567.4	14914.4	534.9	94.3%	26.29	27	15	22	10
ALMONDS	2,974.8	157577.6	2,767.6	93.0%	52.97	14	7	13	6
PEACHES_AND_NECTARINES	6,774.0	411,365.7	6,276.3	92.7%	60.73	8	3	7	5
PLUMS	2,669.4	78924.0	2,367.2	88.7%	29.57	15	8	14	9
Melons. Squash and Cuccumbers	2,489.3	7,638.9	2,161.0	86.8%	3.07	16	19	15	24
SAFFLOWER	47,619.8	67595.7	40,198.6	84.4%	1.42	3	9	3	27
PISTACHIOS	6,082.8	61538.6	4,550.2	74.8%	10.12	9	10	11	13
Misc Deciduous	692.9	2,366.0	517.9	74.7%	3.41	24	23	23	23
VINEYARDS	4,770.2	182,519.5	3,502.5	73.4%	38.26	11	6	12	7
ALFALFA_&_ALFALFA_MIXTURES	57,974.2	221755.4	41,783.9	72.1%	3.83	2	5	2	21
Onions & Garlic	2,414.0	10,997.2	1,720.2	71.3%	4.56	17	18	16	19
Olives	945.7	7045.4	672.0	71.1%	7.45	20	20	19	16
WALNUTS	6,982.2	52202.6	4,560.9	65.3%	7.48	7	11	10	15
Asparagus	959.7	1618.8	576.0	60.0%	1.69	19	26	20	26
Dry Beans (all type)	392.7	3,058.2	226.0	57.5%	7.79	30	22	29	14
CORN	24,336.7	30,536.0	9,564.9	39.3%	1.25	4	13	6	28
Flowers, Nursery & Xmas Trees	569.2	13,893.6	217.4	38.2%	24.41	26	16	30	11
Pears	104.8	1314.5	34.9	33.3%	12.54	38	28	34	12
Jojoba	41.2	10.2	10.0	24.3%	0.25	41	37	37	29
Misc Truck Crop	646.8	2,248.5	22.2	3.4%	3.48	25	24	36	22
Carrots		32557.0	242.0	--		#N/A	12	28	#N/A
WHEAT		13392.7	13,801.2	--		#N/A	17	4	#N/A
Oats		1486.7	941.5	--		#N/A	27	18	#N/A
Broccoli		207.9	574.6	--		#N/A	32	21	#N/A
Grain Sorghum		189.9	42.3	--		#N/A	33	33	#N/A
Cauliflower		184.8	121.0	--		#N/A	34	31	#N/A
Misc. and mixed grain and hay		102.3	281.0	--		#N/A	35	26	#N/A
Prunes		102.2	26.0	--		#N/A	36	35	#N/A
Cabbage		9.3	5.0	--		#N/A	38	39	#N/A
Peppers (Chilli, Bell, etc.)		7.5	7.5	--		#N/A	39	38	#N/A
URBAN/RESIDENTIAL	14,163.8			0		5	#N/A	#N/A	
Dairies	3,993.1			0		12	#N/A	#N/A	
Pasture, General	3,726.1			0		13	#N/A	#N/A	
Railroads	1,850.7			0		18	#N/A	#N/A	
Poultry Farms	857.4			0		21	#N/A	#N/A	
Sudan	856.8			0		22	#N/A	#N/A	
Airport	798.4			0		23	#N/A	#N/A	
Commercial, Misc	459.3			0		28	#N/A	#N/A	
Industrial, Misc	410.3			0		29	#N/A	#N/A	
Turf, General	254.0			0		33	#N/A	#N/A	
Peas, General	203.6			0		34	#N/A	#N/A	
Celery, General	109.4			0		37	#N/A	#N/A	
Sweet Potatoes	101.9			0		39	#N/A	#N/A	
Oranges, general	60.1			0		40	#N/A	#N/A	
Figs	16.3			0		43	#N/A	#N/A	
Idle Land	4.2			0		45	#N/A	#N/A	

Another reason for lower pesticide-applied acres from PUR compared with the CDWR may be the lack of pesticide applications for certain individual crops. We found that organic farming practices do not contribute significantly to cropped area in Kings County[2], although some types of crops may not have pesticides applied for other economic reasons. An additional reason for discrepancies may be errors in data collection that are difficult to assess, such as data entry faults and errors when data is aggregated to the section level.

From 1991 to 1994, the PUR records pesticide applications throughout the year.[3] Similar PUR data for the period 1985-1990 is also available for restricted pesticides. Future epidemiologic studies can utilize this continuous exposure information in the study of health outcomes that have latency periods between exposure and clinical manifestation of the disease.

Conclusions and Future Research

In general, since most cropped lands have pesticide-applied acres from PUR that are less than the number of "true" acres cropped from CDWR, we can assume that the PUR is a reasonable assessment of the agricultural pesticide application for Kings County in 1991. The seasonality of crops and the absence of pesticide applications to certain crops and fields may explain the differences that exist.

Future research involving the PUR should consider several issues. First, the results presented here are for a single county. This technique should be applied to other counties in order to strengthen these findings. Additionally, future research may examine data at the higher spatial resolution of the section level, may provide additional insights into the ways in which pesticide applications are distributed across the landscape. For example, in Table 2 the CDWR and PUR acreages for crops were compared at a countywide level. This same type of comparison could be done for selected sections, to see if patterns exist for the way in which these different databases represent pesticide applications. Examination at this resolution may also lead to a better understanding of multiple cropping to the same field.

[2] Although we could not obtain a quantified estimate of organic acreage by county due to confidentiality issues, communication with state pesticide regulatory agencies indicated that the organic acreage in Kings County is insignificant.

[3] 1995 and 1996 PUR is anticipated to be available by late summer 1999.

VII. OBJECTIVE 2: ASSESSMENT OF THE GAP DATA SET USING CDWR

Introduction

The purpose of this analysis was to determine to what extent the GAP data set might be useful in locating agricultural crops with the goal of refining the spatial resolution of the PUR data set. Since CDWR data are not currently consistently available statewide, the statewide GAP data set was identified for potential use in epidemiologic research that may be performed at that extent. GAP was compared with the California Department of Water Resource (CDWR) data set, using CDWR as the "gold standard". A description of the CDWR data set is presented in the previous section.

GAP data are part of a nationwide program to assess habitat suitability for different wildlife species (Scott and Jennings 1997). Analysis for California is complete, and was conducted by the United States Fish and Wildlife Service (USFWS) and the University of California, Santa Barbara. The GAP land use classification system is based on a modified Holland and Anderson classification scheme (Anderson et al. 1976, Holland 1986; Davis, 1998). GAP classifies 42 general land-use/land-cover types for the purpose identifying critical habitat types. The primary spatial data source for the GAP was Landsat Thematic Mapper data from the early 1990s obtained at a pixel resolution of approximately 30 m^2. In contrast, the CDWR data are based on 100% ground truthed field level data with a minimum mapping unit (MMU) of 2 acres, classifying over 80 different crop species. CDWR data used in this analysis were for the years 1988 (San Joaquin) and 1991 (Kings).

Methods

Before analysis could be performed, CDWR and GAP were mapped to a common classification scheme. The CDWR data set was reclassified to simulate the classification system of agricultural cover types used by GAP. The resultant classification scheme identified 12 basic land use practices, with an emphasis on agricultural land. The resultant scheme classes were: agricultural land, row and field crop, grain crop, pasture, mixed barren land, eucalyptus and orchard/vineyard, native vegetation, riparian areas, surface water and urban land use.

The two data sets were combined using GIS overlay techniques, resulting in one land-cover data set that contained attributes from each original data set for each polygon. Combining the data allowed identification of areas misclassified by GAP in comparison with CDWR. The results of this overlay procedure are presented in Figures 2 and 3. It should be noted that the original spatial resolutions and intents of these data sets differ – the GAP dataset was developed as a regional level data set (Figure 4), so it would be expected to show less detail and more smoothing than the CDWR data, which was collected field by field. Figure 5 shows the results of a GAP/CDWR overlay in an area in San Joaquin County near the city of Manteca, demonstrating how areas of misclassification may arise due to scaling issues alone. However, even with this scaling caveat in mind, the results of such an overlay analysis is informative as to the overall correspondence between these two data sets.

The classification accuracy of GAP for crops was determined by examining the misclassification error at differing levels of crop aggregation. The different level aggregations of CDWR classes into the GAP agricultural land class are presented in Table 3. Initial classification comparisons performed at Level 1 resulted in poor accuracy for most agricultural land use, due to the diversity of CDWR land classes that were subsets of the type agricultural land. Thus, more specific agricultural land-use practices were aggregated into the general category agricultural land to improve classification accuracy (Levels 2 – 4). GAP used the category agricultural land as a "catch all" category for areas related to agricultural production that could not be refined into categories that are more precise.

TABLE 3. AGGREGATION SCHEME LEVELS FOR INCORPORATING CDWR CLASSES INTO GAP SCHEME.

Level 1	Level 2	Level 3	Level 4
Agricultural Land	Agricultural Land	Agricultural Land	Agricultural Land
Row and Field Crop	-	-	-
Grain Crop	-	-	-
Rice	-	-	-
Mixed Barren Land	-	-	-
Pasture	Pasture	-	-
Orchard / Vineyard	Orchard / Vineyard	Orchard / Vineyard	-
Eucalyptus	Eucalyptus	Eucalyptus	Eucalyptus
Native Vegetation	Native Vegetation	Native Vegetation	Native Vegetation
Riparian Vegetation	Riparian Vegetation	Riparian Vegetation	Riparian Vegetation
Surface Water	Surface Water	Surface Water	Surface Water
Urban	Urban	Urban	Urban

Results

The original classification comparison (Level 1) resulted in very poor accuracy, which is attributed to the large number of acres identified as agricultural land by GAP, but not included in the additional agricultural categories which were part of the CDWR scheme (Table 4). The GAP classification scheme uses agricultural land as a "catch all" category for any land-use activity

Figure 2.

GAP and CDWR (1988) Comparison for San Joaquin County

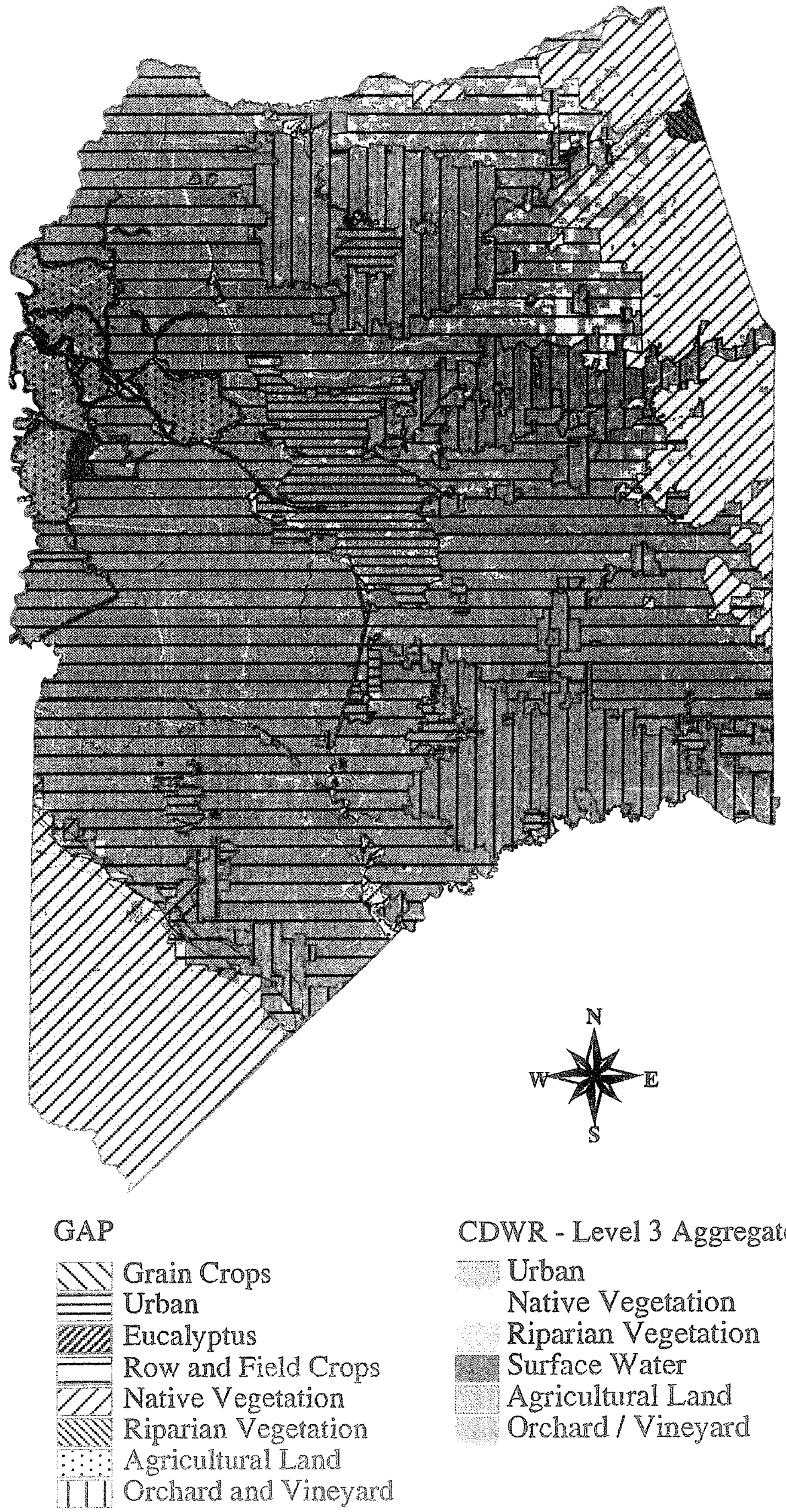
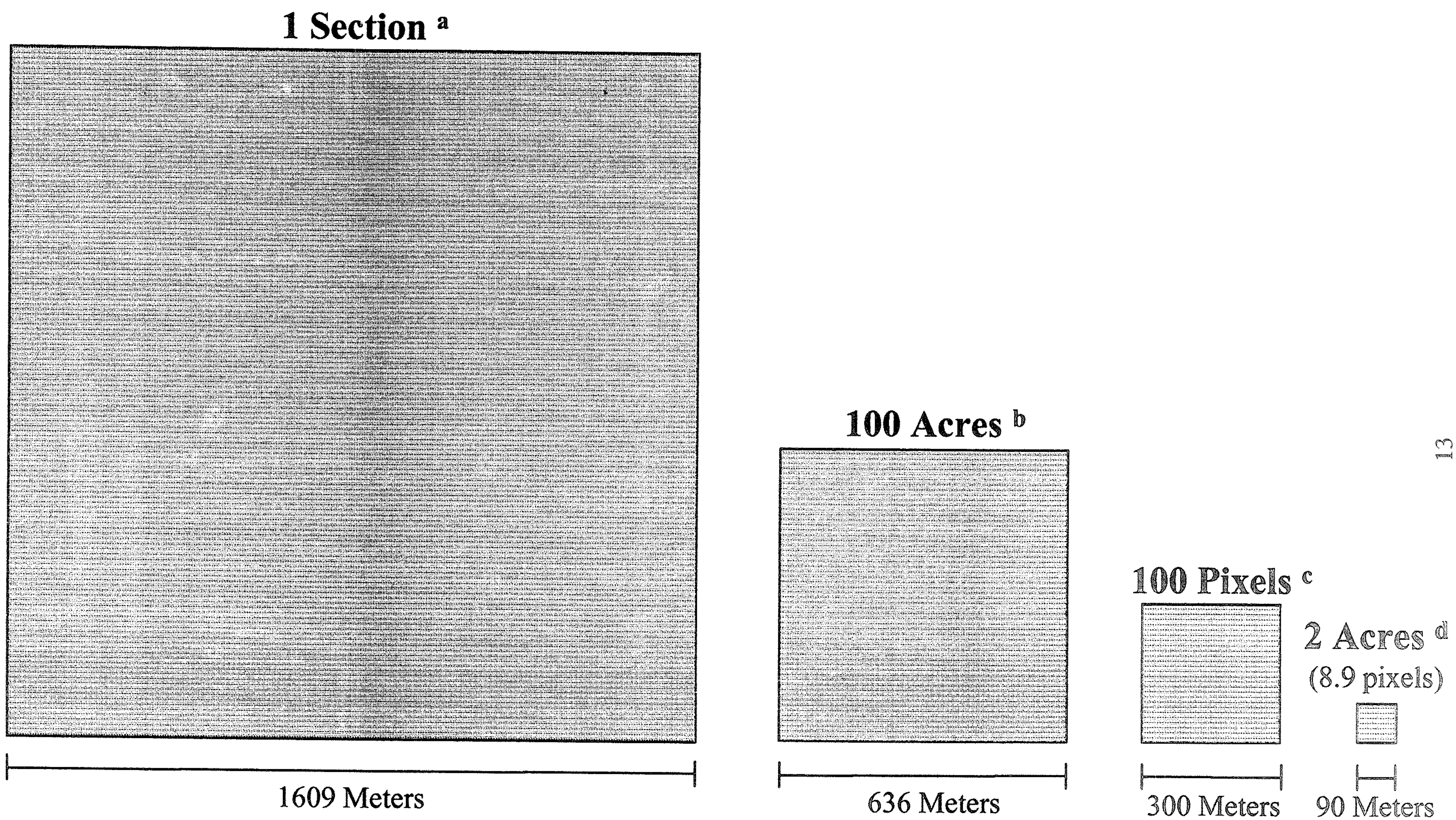


Figure 4. Difference In Scale Between Various Resolution Elements



[a] One section (1 $mile^2$) represents the the resolution element of the PUR data set.

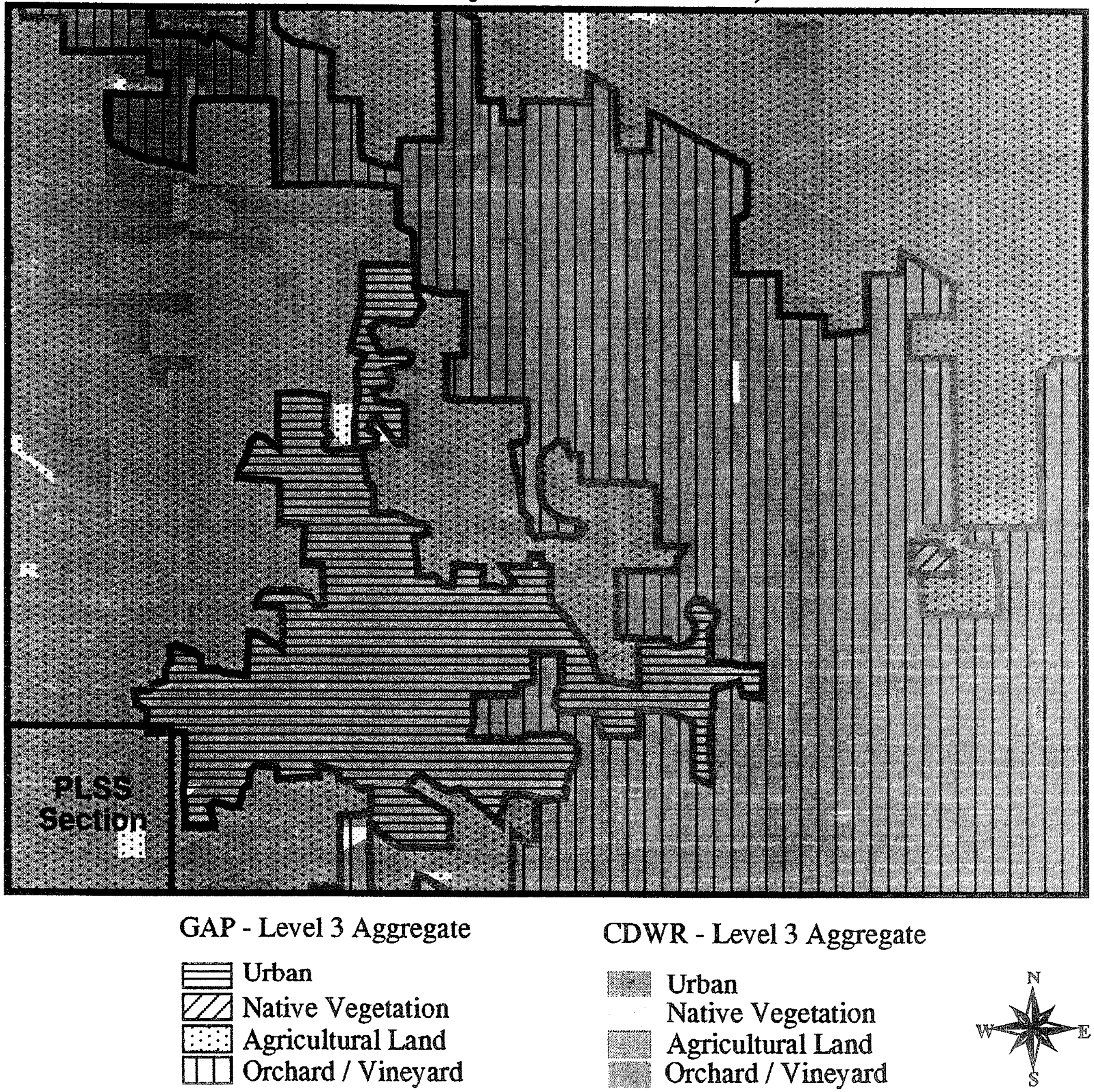
[b] Represents the average area treated with pesticides in a single application in 1988, 1989, and 1991, reported by the PUR data set.

[c] One pixel represents 30 $meters^2$, the original resolution element of the Landsat TM-derived GAP data set.

[d] Represents the minimum mapping unit for the CDWR data set.

Figure 5.

CDWR Overlayed with GAP, Manteca



related to agricultural production which was not spectrally separable during that classification process (including for example farmsteads, feed lots, and cultivation areas). The GAP data analysis demonstrated increasing accuracy as the classifications were aggregated into fewer categories (Levels 2 to 4). Level 3 showed the highest performance for both San Joaquin and Kings counties, as indicated in Table 4.

TABLE 4. CLASSIFICATION ACCURACY OF GAP AT DIFFERENT AGGREGATION LEVELS.

Land Cover Type	San Joaquin County				Kings County			
	Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
Agricultural Land	1.06	61.63	81.27	89.14	1.38	71.36	80.51	84.01
Row and Field Crop	40.45	-	-	-	12.82	-	-	-
Grain Crop	0.00	-	-	-	0.00	-	-	-
Rice	0.00	-	-	-	0.00	-	-	-
Mixed Barren Land	100.00	-	-	-	0.00	-	-	-
Pasture	0.00	0.00	-	-	0.00	0.00	-	-
Orchard/Vineyard	78.48	78.48	78.48	-	71.93	71.93	71.93	-
Eucalyptus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Native Vegetation	90.66	90.66	90.66	90.66	94.36	94.36	94.36	94.36
Riparian Vegetation	12.27	12.27	12.27	12.27	0.00	0.00	0.00	0.00
Surface Water	0.00	0.00	0.00	0.00	100.00	100.00	100.00	100.00
Urban	81.74	81.74	81.74	81.74	68.79	68.79	68.79	68.79

Note: All values are reported in percent.

The Level 3 aggregate improved identification of agricultural land in both counties by incorporating grain crops, row and field crops, pasture, mixed barren land and rice into the agricultural land category, while retaining spatial definition of orchards/vineyards. Table 5 shows classification accuracy at the Level 3 aggregation for San Joaquin and Kings counties.

TABLE 5. CLASSIFICATION ACCURACY OF GAP AT LEVEL 3 WITH TOTAL ACREAGE COMPARISONS

Land Cover Type	San Joaquin County		Kings County	
	Level 3 Accuracy[1]	Deviation of Total Acres From CDWR[2]	Level 3 Accuracy[1]	Deviation of Total Acres From CDWR[2]
Agricultural Land	81.27	+10.03	80.51	+21.05
Eucalyptus	0.00	-179.42	0.00	0.00
Native Vegetation	90.66	-12.06	94.36	-30.61
Orchard/Vineyard	78.48	-5.08	71.93	-65.5
Riparian Vegetation	12.27	+306.34	0.00	0.00
Surface Water	0.00	0.00	100.00	-74.17
Urban	81.74	-9.71	68.79	-36.82

1 Aggregation of rice, grain crop, row and field crop, pasture, and mixed barren land into agricultural land.
2 Reported in percent deviation of acres using CDWR as the baseline.

The results show that GAP overestimated the total acres in the class agricultural land in both counties (San Joaquin ~10% and Kings ~21%) and the class riparian vegetation in San Joaquin county (~306%). GAP underestimated the total acres in all other cover classes. Accuracy for agricultural land at this aggregation level was good, achieving approximately 81% identification accuracy for both counties, while orchards/vineyards achieved fair identification accuracy - 78% for San Joaquin county and 72% for Kings county.

Discussion

Much of the classification error associated with the GAP data may be attributed to differences in the way that the CDWR and GAP data sets were collected and processed, and the purposes for which they were originally developed. In this research, we have put the data to uses somewhat different than for what they were developed. An awareness of the way in which the data sets were created adds insight to the discussion of results for this assessment. The GAP data was developed from classification of Landsat TM data (a grid-based GIS raster data set), which is collected by measuring the spectral reflectance of an area approximately 30 m^2 (.2 acres). Reflectance values of spectral bands in both visible and infrared ranges are used to classify this remotely sensed data into cover types or land-use. The classification is based on areas of "like" spectral reflectance, i.e. areas that "look" alike spectrally will be classified as the same land use or cover type. Classification is determined by the predominant cover type or land-use type, for example, different areas of predominantly orchards and vineyards would likely be classed alike (assuming they are spectrally recognizable), despite intermixing of other land uses within their extents. The classification process groups areas containing pixels with similar reflectance, so that the effective resolution of ground elements on the data set is not effectively the individual pixel level, but at the level of aggregated groups of pixels. The GAP data set is designed for use at a scale of 1:100,000. The GAP data is thus a more spatially generalized data set than the CDWR data, since the homogeneous polygons classified in GAP are larger than the 2-acre minimum mapping unit of the CDWR data. In addition, the GAP attribute classification scheme grouped much of the agricultural land classes together, while the CDWR attribute classification scheme identified many separate crop classes. Figure 5 depicts the way in which GAP represents complex land-use activities in a manner different than the CDWR data set.

The CDWR spatial data set was developed from aerial photography, which was used to define polygons of homogeneous land use (GIS vector data). These polygon maps were then visually checked in the field to assign a land use to each polygon. This resulted in a relatively complex data set, which defines land use with a high degree of spatial and attributed accuracy and resolution. When this data was compared with the lower resolution GAP data set, a large amount of misclassification resulted at finer classification levels (Levels 1 and 2). Thus, much of this misclassification can be directly attributed to the manner in which the two data sets were created.

Despite these differences, GAP does a reasonably good job of delineating agricultural land and orchards/vineyards on the landscape, in a manner that is amenable to use in future epidemiological work. Referring again to Figures 2 and 3, which depict the Level 3 aggregate of the CDWR data over laid with the GAP data set, GAP identified boundaries of predominant land use categories with a high degree of accuracy. Thus, this data set may prove useful in spatially refining areas of pesticide use contained in the PUR data, which has a resolution of 1 mi^2, enabling researchers to more accurately delineate areas of pesticide application.

Conclusions and Future Research

Based on our analysis of San Joaquin and Kings counties, we feel that the GAP data set can be used with reasonable accuracy and confidence to identify locations of agricultural land and orchard/vineyards. Both counties responded similarly under analysis, leading to the conclusion that the data set may be applied to other agricultural areas with similar crop types. Further analysis needs to be performed in other agricultural areas of California to determine how GAP can be used to identify agricultural land in areas where crop species and cultivation practices are different then those in San Joaquin and Kings county.

VIII. OBJECTIVE 3: IMPROVING THE SPATIAL RESOLUTION OF THE PUR DATA SET

Introduction

As noted above, the Pesticide Use Reporting database is a collection of pesticide applications reported at the spatial resolution of the section in the Public Lands Survey System. The area of a PLSS is normally 640 acres, or approximately 1 $mile^2$. This coarse resolution information is of limited utility for exposure assessment by epidemiologists. Via this objective, we present results from a study of methods for improving the spatial resolution of the PUR data set beyond the section level. We employed two approaches: (1) examining the availability of the original maps from which the PUR was derived and using a field-identifier from those maps to create a linkage between spatial and attribute data, and (2) assessing the viability of performing a merge between the PUR and another spatial database such as GAP.

Field-identifier linkage

We contacted Kings and San Joaquin County Agricultural Commissioners about the availability of PUR data at a resolution greater than the section level, which could be used to delineate the boundaries of fields within each section where pesticide applications had been recorded. Such original hard-copy data is not available. The counties we studied do not maintain any historical records about past fields identified in the PUR; i.e. the maps are not recorded and archived in any systematic fashion. In addition, the "unique" field-identifier given to a landowner is randomly generated on a yearly basis, so it is not possible to infer the distribution of fields within the county that may vary over time.

Even if other counties maintained such data, a number of problems with their use should be anticipated. First, these maps may not be geographically referenced, which would be necessary for integration with the location of participants in a case control study using GIS. Second, the cost of digitization into a GIS format would be an expensive undertaking. A single county with ~5,000 different fields may take upwards of 40 hours to assimilate into a GIS. Considering the need for wide-scale application, digitizing 10 or 20 counties would be extremely costly. Another problem would be ascertaining the spatial accuracy of field boundary locations on hand-drawn maps or those derived from aerial photography. When the cost of the processing the original maps and the effort involved in processing such maps is considered, the viability of using the original PUR hard copy maps for this method seems very low at this time. In the event that counties went to a digital GIS-based method of collecting and maintaining consistent geographic PUR Field-id information, the viability of using such maps would greatly increase.

Spatial Merge

Assumptions

A second option for improving PUR's spatial resolution was examined, merging PUR with an existing spatial database. There are a number of assumptions that must be made for this process. The records in PUR must be accurate and inclusive of all pesticide applications of interest, which may depend on the health outcome of interest and its relationship with particular pesticides or pesticide classes. Additionally, the relationship between the spatial database's land cover classes and the PUR classification must be precisely mapped, i.e. there must be compatible categories in both data sets (as in the reclassification of codes in Objectives 1 and 2). Finally, the spatial data must be representative of the actual landscape for which pesticide applications recorded in the PUR are recorded. This is particularly important in regions where crops are rotated frequently, either seasonally or year-to-year.

Limitations and Benefits

There are certain limitations and benefits of performing a spatial merge between PUR data and spatial data that refines the pesticide application locations. The benefit gained from the merge is related to the level of classification resolution of the spatial database, both the number of classes and the minimum mapping unit of the data. Data at a spatial resolution finer than the section level could provide refinement (beyond the PLSS section level of the PUR data) in the assessment of exposure across a region. One of the limitations is in finding spatial data of good quality that is complete across a sufficiently large geographic area to allow the collection of enough cases for epidemiologic studies. In California, depending on the outcome prevalence and other factors, this may require a statewide data set.

Spatial Data Selection

Of the land use/land cover spatial data sets considered for this aspect of the study, which are listed in Table 6, the GAP data set was deemed the most appropriate for merging with the PUR. GAP covers the entire state, and has been shown to be reasonably accurate in locating coarse land use classes such as 'Agricultural Lands' or 'Orchards/Vineyards' (see Objective 2, above). The CDWR data set would provide more detailed land use/land cover information for a merge, but it is currently limited in spatial and temporal availability. The USGS LULC (Land Use-Land Cover) data set is not appropriate for this portion of the analysis because of its temporal characteristics – it covers the mid-1970's. The Multi-Resolution Land Classification - National Land Cover Database (MRLC-NLCD) data set is similar to GAP in its spatial resolution level and temporal coverage for the early 1990's, but is not available for California at this time (however it is anticipated to be available late in 1999). Table 6 outlines characteristics of these spatial data sets relevant to improving the spatial resolution of the PUR via a spatial merge.

TABLE 6. AVAILABLE SPATIAL DATA SETS WITH POTENTIAL FOR MERGING WITH PUR.

Data set	Number of land use/land cover classes	Comments	Source
CDWR	~83	Most ideal for merge + Detailed land cover classes + 100% Accuracy + High spatial resolution (2 Ac MMU) - Limited spatial and temporal coverage	California Department of Water Resources
GAP	~12	Somewhat ideal for state wide-coverage + State-wide coverage + Reasonable accuracy at a coarse grouping of thematic classes - Requires major aggregation of PUR's classification - Only one time point available	CA GAP Analysis Project. Biogeography Laboratory, UC Santa Barbara
LULC	~12	Not appropriate for merging with post- 1990 PUR - Timeframe does not overlap PUR's collection dates (collected in mid 70's) + State-wide coverage	US EPA conversion of US Geological Survey files
MRLC	15	Potential source for state-wide analysis + State wide coverage - Unknown release date for California See metadata for MRLC (http://nsdi.epa.gov/nsdi/projects/r3_mrlc.html#section1)	U.S. Geological Survey, EROS Data Center

Methods

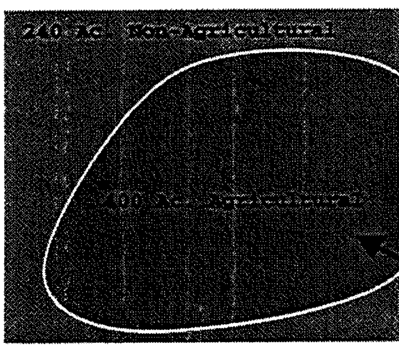
The process of merging PUR with a spatial database is straightforward. This is best explained by examining a sample PLSS section, such as is seen in Figure 6. Assume that in this section we know that the agricultural pesticide application will only occur within agricultural lands that are identified by the spatial database. Without the spatial database, i.e. using PUR alone, we only know that the pesticide application occurs somewhere in the section, and would often assume that it is evenly distributed across the section. However, using the increased knowledge of location of agricultural lands within a section can refine assumptions of where pesticides are being applied. This linkage of pesticide application information with the spatial database may make a significant difference in the determination of exposure for a household.

To demonstrate, the hypothetical situation in Figure 6 is used as an example. In the first record of the PUR, atrazine was applied to 200 acres of corn. We then assume that out of the 400 acres in agricultural land in the section, 200 acres had atrazine applied. This leads to a 50%

probability of exposure within this agricultural area. Similarly, the propachlor application could be shown to have a 12.5% probability in this area. This process could be repeated for all pesticide applications and for every section in the study area.

If we did not use the refined spatial data, we could only assume equal probability of application across the entire section. For the August 3 atrazine example, this would result in a 31.35% (200 acre application divided into 640 acres of the section) probability of exposure within the entire section. For a residence within the agricultural area, this would likely be an underestimation of potential exposure, whereas for residences in the non-agricultural portion of the section this would likely be an overestimation.

Figure 6. Example - Merging Spatial Data with PUR



After merge of PUR with a spatial database we know more precisely where the exposure is occurring.

PUR EXAMPLE

Section ID	Field ID	Crop Applied	Chemical	Date Applied	Application Amount (tot. lbs.)	Acres Applied		Probability of Exposure in Agricultural areas
001	3055	Corn	Atrazine	August 3, 1991	35	200	400	50%
001	3055	Corn	Atrazine	May 2, 1991	26	250	400	62.5%
001	3055	Corn	Propachlor	April 2, 1991	2	50	400	12.5%
001	489F	Barley	DDT	March 3, 1991	3.5	100	400	20%

We performed a preliminary analysis in Kings County using 1991 PUR data to judge the effectiveness of merging the PUR and GAP for increasing the spatial resolution of exposure assessment. A GIS overlay function joined the GAP and PLSS section boundaries for Kings County. Adding the spatial information from GAP to the PUR via this method would improve exposure classification in 34% of the sections in the county. That is, the increased resolution of the location of known agricultural areas was made available beyond the section level in 34% of the sections. Merging with the higher spatial resolution CDWR data set for the same area would improve exposure classification based on agricultural lands in 76% of the sections. This variation results from the difference in spatial resolution of the two data sets, the CDWR being a higher resolution data set.

Assigning Exposure to Households

Once the spatial merge process has been performed throughout the study area, a protocol can be established for assigning actual pesticide exposure to particular households. Four potential methods of refining exposure probability at a point residence are illustrated in Figure 7.

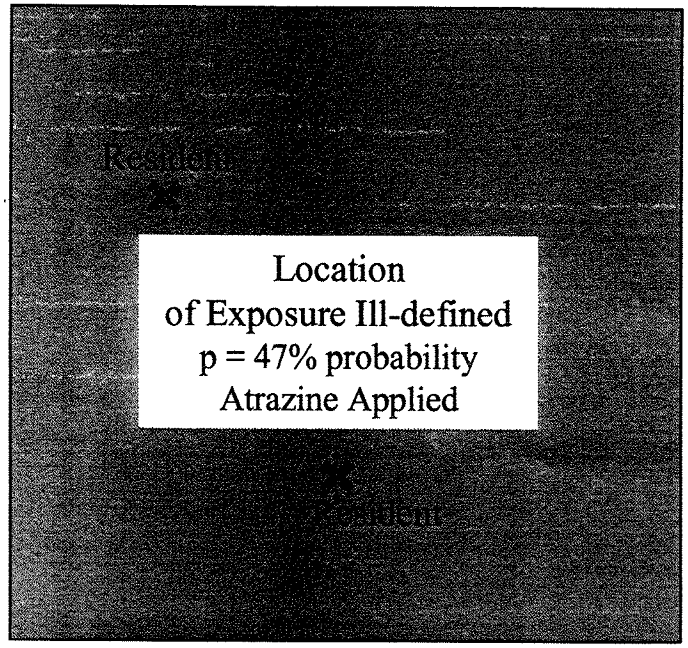
Tile A of Figure 7 illustrates the exposure calculation when exposure is measured at the PLSS section level using PUR. In this example, there is a known atrazine application to 300 acres in the section. Probability of exposure is assigned equally to all residence locations, using 300 acres/640 acres (area of a section) which equals 47%. Thus, an equal probability of exposure is assigned to all residents regardless of proximity to agricultural crops or pesticide application areas. At this resolution it is difficult to develop exposure risk assignments for a particular residence.

Tile B shows a simple point-in-polygon exposure, whereby a household is assigned the exposure probability of the polygon in which it is found. This approach refines the estimation of probability of pesticide application, since an agricultural area's location and dimensions are known. Thus, 300 acres of atrazine application can be distributed over the 400 acres of agricultural land, giving a 75% exposure probability for residences within the agricultural polygon. With respect to point residence locations, the refinement would vary with the resolution of the spatial database used. One difficulty with this approach is that it ignores the fact that pesticide applications are seldom confined to the area of application--chemicals may be spread by the air or carried through the ground and contaminate groundwater in wells.

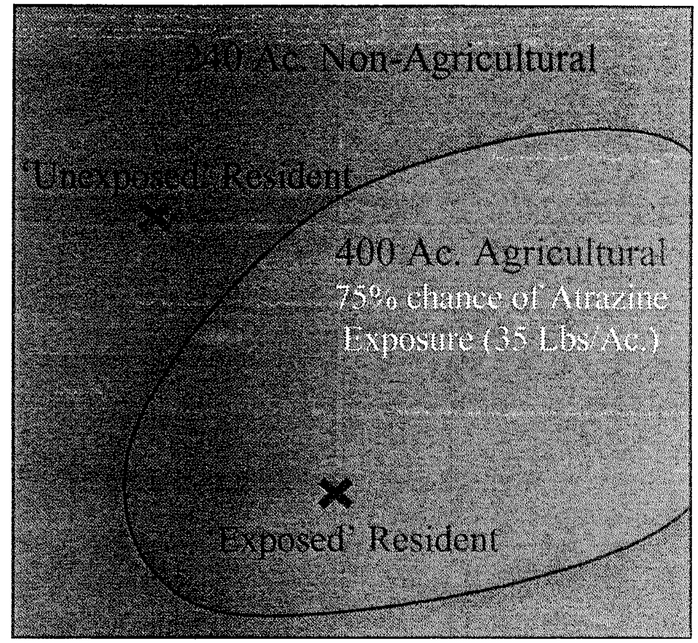
Tile C illustrates the incorporation of a buffer distance around a residence. Buffers may be used to calculate exposure based on proportional overlap of exposed zones. In estimating exposure occurring from drifting pesticides, a probable drift distance based on studies of pesticide dispersion in the environment could be used to generate buffers (Ward et al., 1999). This method does not account for transport of pesticide from sources lying outside of the buffer. If such transport did occur it would result in an additional potential for exposure.

Figure 7. Potential methods for refining exposure probability.

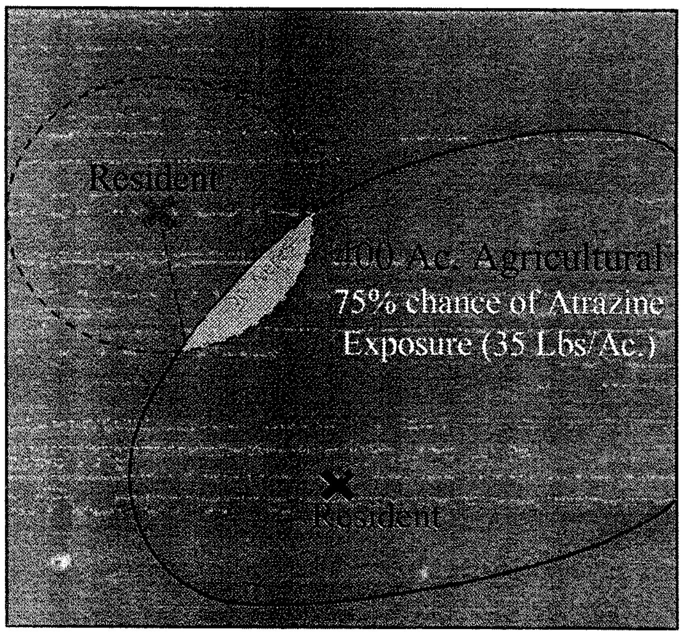
A. PUR Only



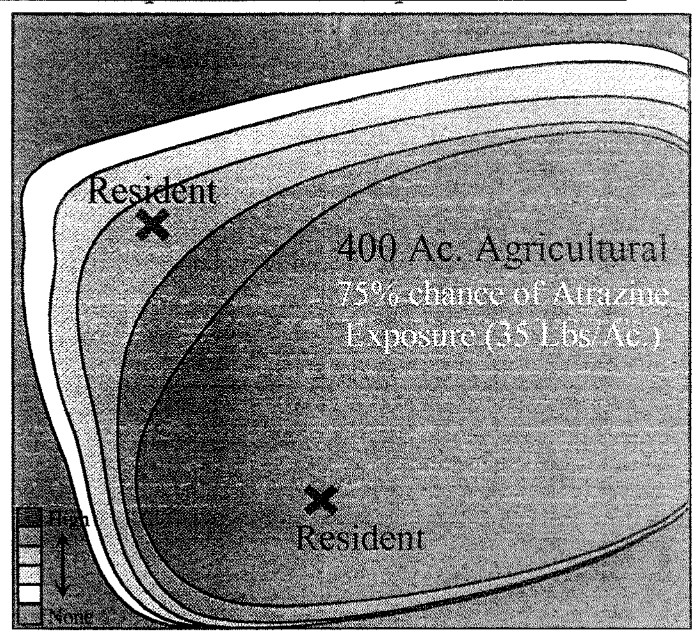
B. PUR and Spatial Data: Boolean



C. PUR and Spatial Data: Buffer Proximity



D. PUR and Spatial Data: Dispersion Model



Tile D illustrates the incorporation of a dispersion model, which could be wind or groundwater based. While such modeling may be complex to implement, this method can refine exposure assessment further by incorporating environmental factors that determine pesticide movement. Pesticide deposition may not be uniform across an application area, as it may be influenced by factors such as wind speed and direction, soil type, application type, application medium, time of application, temperature, and humidity. Thus, the region surrounding an agricultural area may be zoned based on modeled probability of higher or lower exposure. The selection of a model for generating such areas would depend on the pesticide of interest; particularly its application method and on the available of data concerning the geophysical factors that govern transport in a particular environmental setting.

Discussion of Results

This preliminary work demonstrates the technical feasibility of using a higher spatial resolution data set to enhance PUR section level data. However, there are a number of refinements necessary before the full-scale application of spatial merge methods could be implemented at a statewide or similar level. The most appropriate statewide database for the spatial merge would need to be identified, depending on data quality, spatial resolution, time points of interest, and attribute classification levels. The MRLC database that has been discussed (see Table 6) is an example of such a potential data source, however the CDWR data on a countywide level has higher spatial and crop type resolution, if the appropriate time frames for a study was available.

Foremost, further evaluation of the PUR is needed, as is indicted in the discussion concerning Objective 1. Multiple applications to the same field need to be accounted for, which is difficult when the appropriate field identifiers are not present in the database.

Preliminary analysis shows that about 34% of the township and range sections in Kings County for 1991 would see an improvement in the resolution of exposure classification when a spatial merge with GAP is utilized. Most of these sections are located near urban and rural regions, which is where much of the population is located in Kings County. If the majority of cases and controls reside in these areas, then using GAP in the spatial merge would be of benefit to an epidemiologic investigation. Otherwise, in comparison to section-based methods, it may not provide significant advantages from the point of view of numbers of subjects whose classification was refined.

Future Research

Applying the spatial merge technique to a county, or a smaller area would allow further details of the above-described process to be. Here the GAP and CDWR were merged via overlay, to determine that increased exposure resolution would benefit in 34% and 76% of sections using GAP and CDWR, respectively. However, further study could elucidate more fully the types of land uses that would be spatially refined, as well as their typical location in the landscape (e.g. peri-urban areas). Since a complete county-level overlay would be quite analytically extensive, even studying a subset of sections of a county would greatly assist in the development of a fully developed exposure assignment protocol, by demonstrating the utility of higher spatial resolution exposure information.

IX. OBJECTIVE 4: IDENTIFICATION OF RESIDENCES CONSTRUCTED ON FORMER AGRICULTURAL LAND USING GAP AND LULC

Introduction

A variable of importance to epidemiological analyses is whether cases or controls reside on former agricultural land, since agricultural chemical residuals in locations that are currently developed for residential use may influence disease outcomes over the time frames that are of interest in many epidemiologic studies.

Methods

Databases for Kings and San Joaquin counties used in the pilot project were evaluated for their possible use in determining whether a subject resides on former agricultural land. Table 7 lists the databases utilized. They were (for both counties): CDWR data, reselected for urban land only; GAP, reselected for urban land only; LULC, reselected for agricultural land only; and a dummy point data set generated using random number techniques to simulate possible subject locations. Table 7 also shows dates and spatial resolution of data sets. Using the three spatial land use/land cover data sets available (CDWR, LULC and GAP), we looked at the earliest time point available for agricultural land (LULC) and how it corresponded with later time points for urban

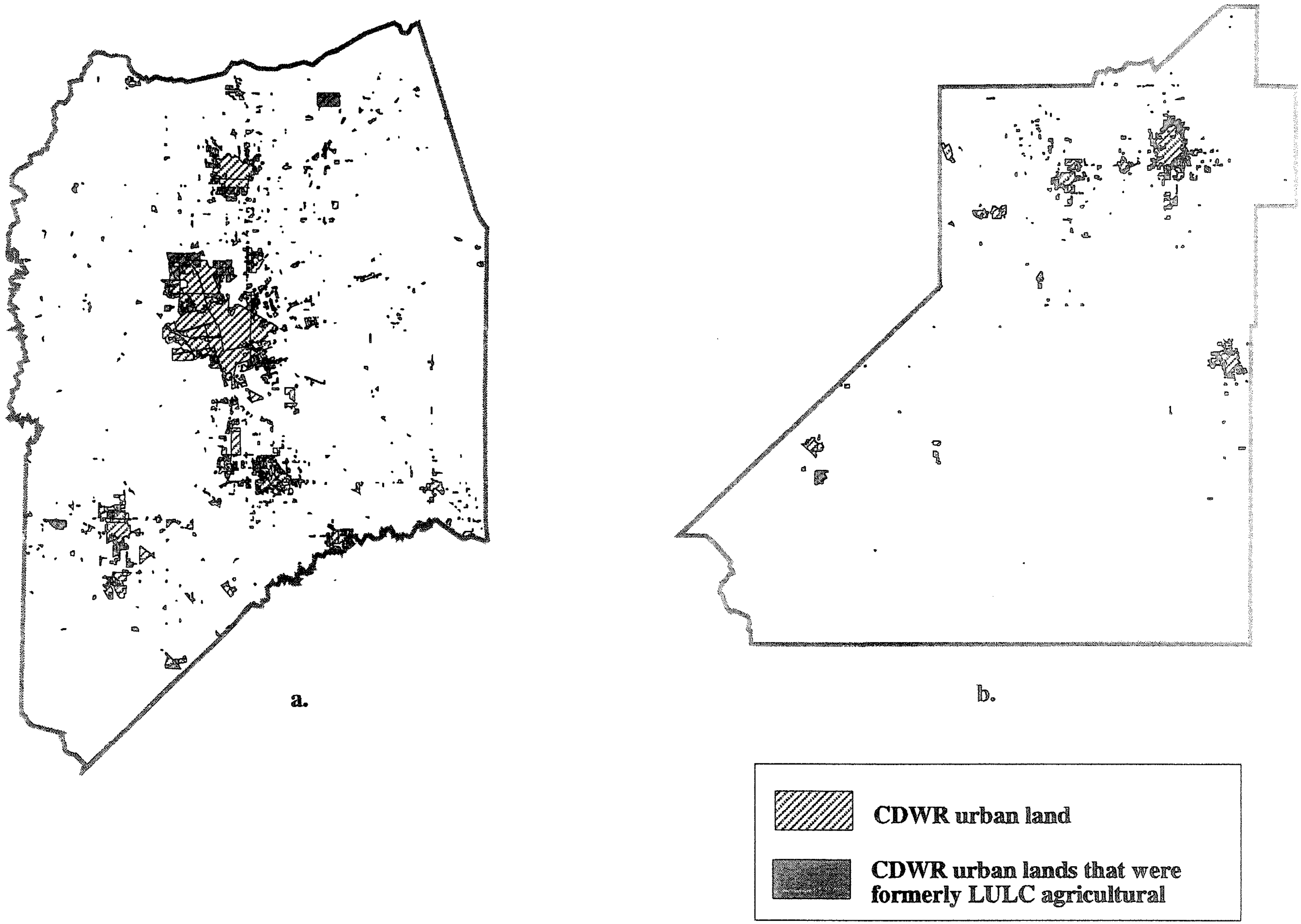


Figure 8. Urban areas in CDWR data sets that were formerly agricultural, as designated by LULC. Dates of CDWR differ by county (see Table 1.) a. San Joaquin County b. Kings County

/data13/calif/projects/formerag.apr/figure1

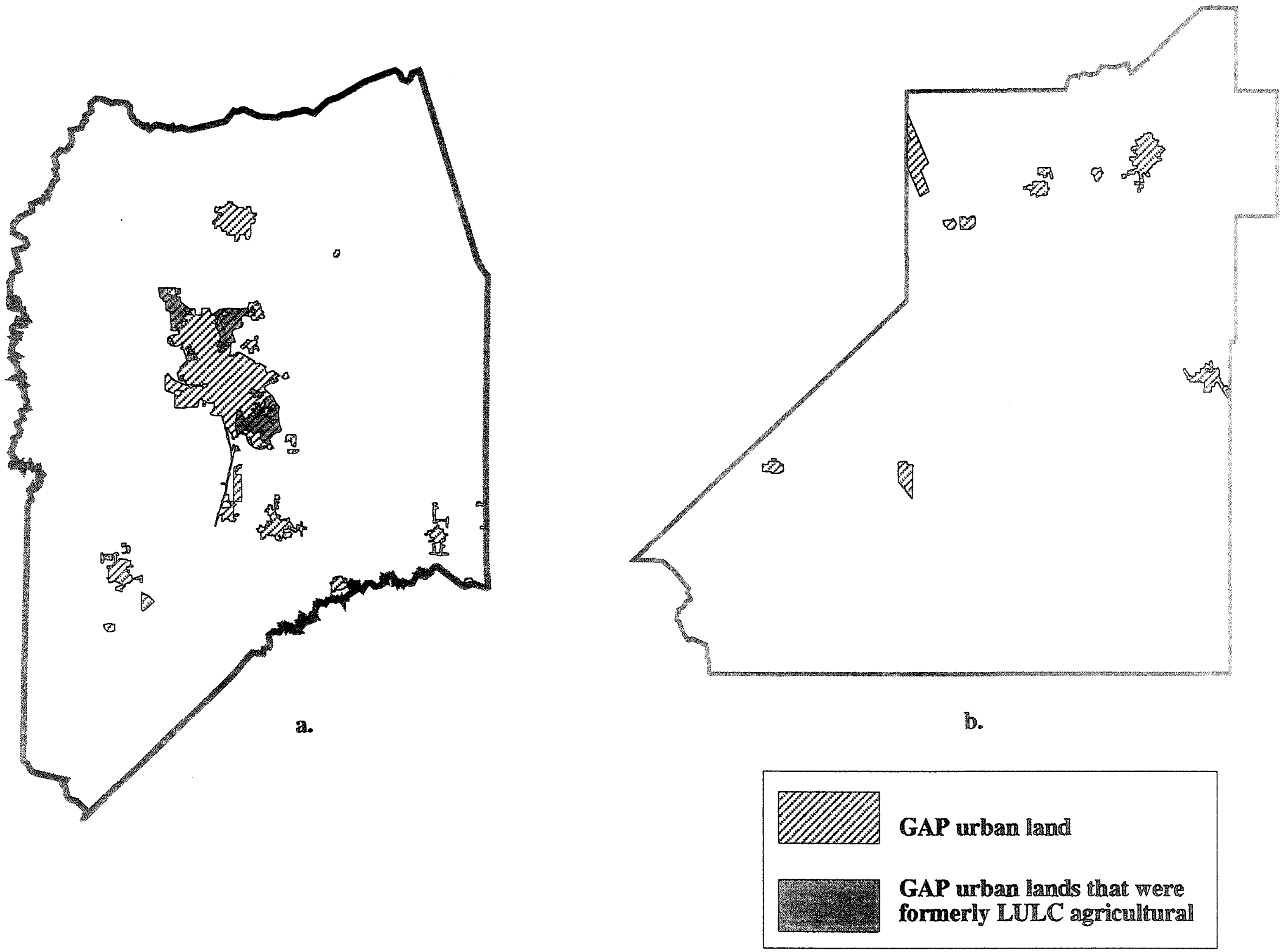


Figure 9. Urban areas in GAP data sets that were fomerly agricultural, as designated by LULC. See Table 1 for dates.
a. San Joaquin County b. Kings County

/data13/calif/projects/formerag.apr/figure2

land. The technique for doing this was a polygon-on-polygon overlay. This demonstrates the location of land areas which are urban land at the later time points, but are also former agricultural land (Figures 8 and 9). We also overlayed point data (randomly generated pseudo-subject point locations) with agricultural polygons (LULC) to demonstrate the utility of point-in-polygon overlay techniques for generating statistics for residence on former agricultural land as part of potential future studies. This type of technique is useful since not all of the persons residing on former agricultural land might necessarily be in urban areas. However, it is often the case that suburban encroachment onto former agricultural lands creates urban "rings" around central urban cores. These suburban rings often contain relatively recently built subdivisions and residences built on former agricultural lands.

TABLE 7. DATABASES USED FOR FORMER AGRICULTURAL LAND DETERMINATION.

Data Set	Date	Reselection criteria	Resolution
CDWR – San Joaquin	1988	Urban (Class 1 U*)	2 acre MMU
CDWR -- Kings	1991	Urban (Class 1 U*)	2 acre MMU
GAP – San Joaquin and Kings	1990-95	Urban (Comp-code 6)	30 meters
LULC – San Joaquin and Kings	mid 70's	Ag (LUCODEs 21,22,23,24)	10 – 40 acre[1]
Pseudo-subject locations	N/A	N/A	points

[1] MMU was 10 acres for manmade objects and 40 acres for non-urban, non-manmade objects. See (http://edcwww.cr.usgs.gov/glis/hyper/guide/1_250_lulc.html) for data availability information.

Results

Figures 8a and 8b show resultant maps of overlays between LULC (agricultural) and CDWR (urban) for San Joaquin and Kings. Areas in dark shading are areas that were urban in the CDWR time period, but agricultural in the 1970s when the LULC data were collected. Figures 9a and 9b show similar maps generated for LULC (agricultural) and GAP (urban). See Table 7 for dates. Note how the difference in resolution of the data sets (CDWR vs. GAP) affects the outcome and look of the resultant overlays. Figures 10a and 10b show the utility of point-in-polygon overlay for determining residence on former agricultural land, using LULC agricultural coverages and randomly generated point data. The GIS could be used to generate input data for statistical analyses on these points, such as correlations of disease outcomes with types of crops in the vicinity. Note also that this is a simplified methodology devised as a demonstration of possible approaches. If desired, more complex analyses could also be performed using multiple time points and other land use/land cover types.

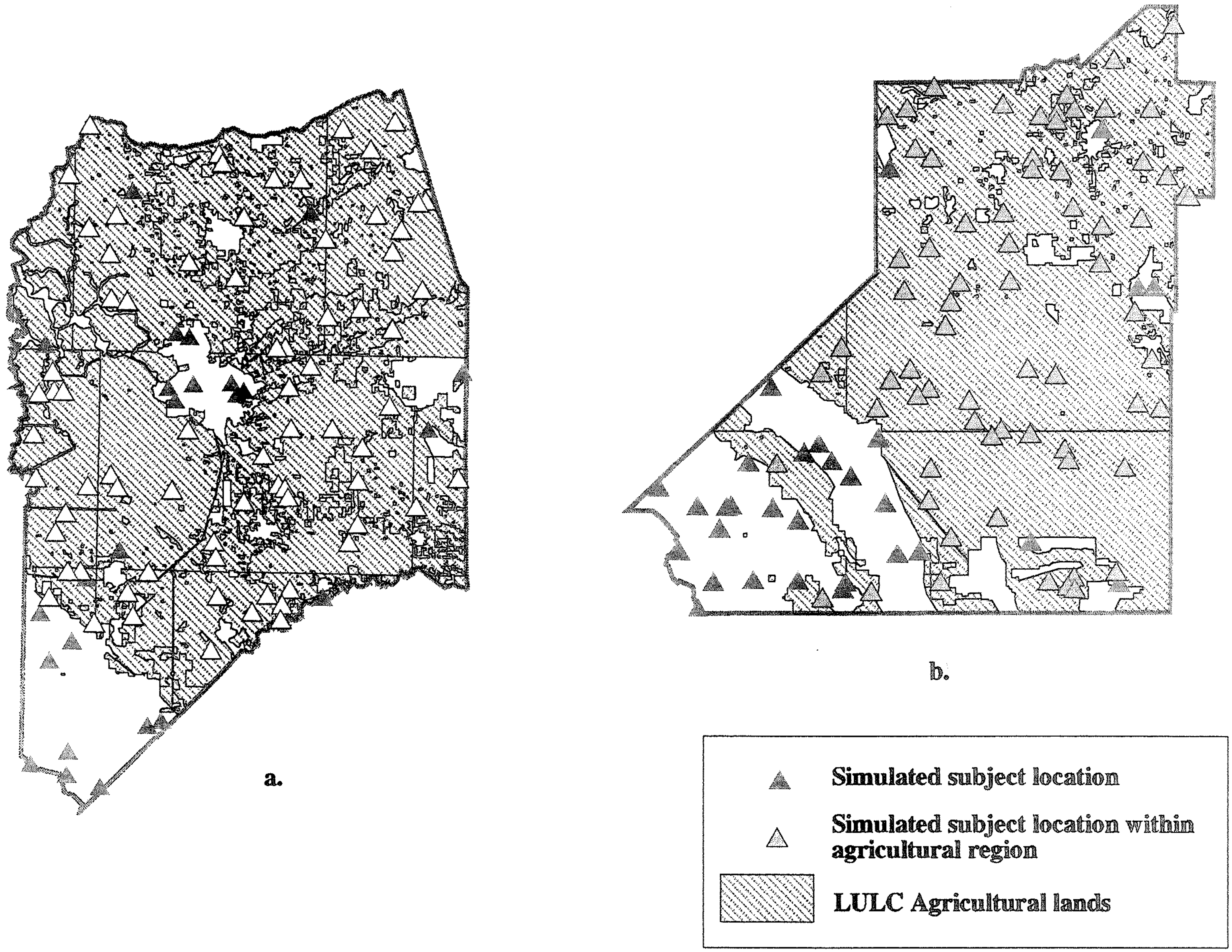


Figure 10. Simulated subject locations point-in-polygon overlay onto LULC agricultural lands.
a. San Joaquin County b. Kings County

/data13/calif/projects/formerag.apr/figure3

Conclusions and Future Research

We have shown that the LULC and GAP/CDWR data sets can be used together to identify lands that were formerly classified as agricultural. This information would be a useful clue in the assessment of historical pesticide application on lands that are now residential. Because of the statewide availability, the ideal data sets for this type of analysis would be the USGS LULC and GAP. Since the LULC data does not have any reported accuracy, we cannot place any specific confidence in this process. Future projects can implement this technique for anecdotal evidence to residual pesticide exposure.

The Multi-Resolution Land Cover (MRLC, another spatial data set briefly mentioned in the 'Increasing the Spatial Resolution of PUR' section) will become available mid to late 1999 for California. This data set is similar to GAP in that it is derived from Landsat TM imagery from the same time period, but will cover the entire nation. This data set could similarly be used in the analysis presented above.

X. OVERALL CONCLUSIONS AND FUTURE WORK

This study has investigated two major issues concerning refinement of pesticide exposures in California. First, we have examined the Pesticide Use Reporting (PUR) database and available spatial data sets that may be used with it to refine exposure to pesticides. Second, we established a procedure that would allow the identification of lands that were formerly agricultural.

To validate the PUR database we compared the acres of pesticide application to a spatial database developed by the California Department of Water Resources (CDWR). The CDWR database is highly detailed (83 informational categories) and highly accurate (100% checked with ground crews), but has only been collected at various time points in 21 counties thus far.

Analyzing Kings County for 1991, we found similarities in reported acres of pesticide application per crop to the acres in the CDWR data, but certain discrepancies between CDWR and PUR acreages warrant future research. In particular, examining the data sets at the section level may prove useful in detailing many of the differences between CDWR and PUR. Additionally, the validation of PUR in other counties for other years would also shed light as to the accuracy of the data.

We validated the GAP data set (using CDWR) to determine its utility for merging with the PUR to refine the location of pesticide exposure. We found that the GAP was reasonably accurate in locating coarse agricultural classes of land use such as 'Agricultural Lands' and 'Orchard/Vineyard' for the two counties and time periods we studied (Kings, 1991 and San Joaquin, 1988).

We analyzed procedures for increasing the spatial resolution of the PUR data set. Both the GAP and the CDWR data sets can be used in conjunction with the PUR for this purpose, but only the GAP is appropriate where wide-scale application is needed, due to availability. In future work, the technique should be applied to an entire county (or a subset thereof) using subject residence point locations to ascertain the utility of using each of the data sets with PUR.

Another issue involved in using refined exposure information is determining the actual exposure to households. Point-in-polygon, buffer and exposure models all can be used for this, but each has a host of its own problems associated with it. Future work may also include the evaluation of these techniques in a small area in conjunction with household dust samples for validation.

We used the USGS Land-Use/Land-Cover (LULC) data set in conjunction with the GAP and the CDWR to identify lands that were classified as agricultural in the mid-1970's. This would assist in the evaluation of household pesticide exposure history for certain types of adverse health outcomes.

XI. REFERENCES

Anderson JR, Hardy EE, Roach JT, Witmer RE. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964, Washington, D.C., (1976).

Beck LR, Rodriguez MH, Legters L. Remote sensing as a landscape epidemiological tool to identify villages at high risk for malaria transmission. Am J Trop Med Hyg 51:271-280, (1994).

Bradman, MA, Harnly, ME, Draper W, Seidel S, Teran S, Wakeman D, and Neutra R. Pesticide exposures to children from California's central valley: Results of a pilot study. J Expos Anal Environ Epidemiol 7:217-234, (1997).

Camann, DE. Investigating the effects of pesticide exposure in the home. Technology Today June:1-7, (1994).

Campbell JB. Introduction to Remote Sensing, New York:Guilford Publications, Inc., (1996).

Davis, F. W., D. M. Stoms, A. D. Hollander, K. A. Thomas, P. A. Stine, D. Odion, M. I. Borchert, J. H. Thorne, M.V. Gray, R. E. Walker, K. Warner, and J. Graae. 1998. The California Gap Analysis Project--Final Report. University of California, Santa Barbara, CA. [http://www.biogeog.ucsb.edu/projects/gap/gap_rep.html]

Glass, G.E., Schwartz, B.S., Morgan, J.M., Johnson, D.T., Noy, P.M., Isreal, E. Environmental risk factors for Lyme disease identified with geographic information systems. Am J Public Health 85:944-948, (1995).

Goolsby DA, Thurman EM, Pomes ML, Meyers MT, Battaglin WA. Herbicides and their metabolites in rainfall: Origin, Transport and deposition patterns across the Midwestern and Northeastern United States, 1990-1991. Environ Science Technol 31:1325- 33, (1997).

Holland RF. Preliminary descriptions of the terrestrial natural communities of California. State of California, The Resources Agency, Nongame Heritage Program, Department of Fish and Game, Sacramento, CA, (1986).

Johnson B, Kamble ST. Pesticide use on major crops in Nebraska-1982, Report No. 10, Lincoln: Environmental Programs, Cooperative Extension Service and Agricultural Research Division, Institute of Agriculture and Natural Resources, University of Nebraska-Lincoln, (1984).

Lam, NS-N and Quattrochi D. On the issues of scale, resolution and fractal analysis in the mapping sciences. Professional Geographer 44:88-98, (1992).

Maas RP, Kucken DJ, Patch SC, Peek BT, van Engelen DL. Pesticides in eastern North Carolina rural supply wells: land use factors and persistence. J Environ Qual 24:426-431, (1995).

Maxwell SK, Nuckols JR, Ward MH. Historical Crop Mapping Using Agriculture Statistics and Landsat Satellite Data, Submitted to Photogrammetry and Remote Sensing. National Agricultural Statistic Service, Agricultural Chemical Usage, http://usda.mannlib.cornell.edu/reports/nassr/other/pcu-bb/

Nuckols, JR, Xiang H, Stallones L, Faidi H. Spatial information technology and analysis in health assessment of rural communities. (Abstract). *Epidemiology* 7(4)S-73 (1996).

Richter ED. Aerial Application and Spray Drift of Antichlolinesterases: Protective Measures. In: Clinical and Experimental Toxicology of Organophosphates and Carbamates (Ballatyne B, Aldrige T, eds). Oxford:Butterworth-Heinemann, Ltd., (1992).

Scott JM and Jennings MD. A description of the National Gap Analysis Program. Biological Division, U.S. Geological Survey, (1997). http://www.gap.uidaho.edu/gap/AboutGAP/GapDescription

Simcox NJ, Fenske RA, Wolz SA, Lee I, Kalman DA. Pesticides in household dust and soil: exposure pathways for children of agricultural families. Environ Health Perspect 103:1126-1134 (1995).

U.S. Department of Health and Human Services. Seventh Annual Report on Carcinogens, Summary 1994. Technical Resources, Inc. (1994).

Ward, MH; JR Nuckols; SJ Weigel; SK Maxwell; KP Cantor; RS Miller. Estimating environmental exposure to agricultural pesticides using remote sensing and a geographic information system: Results from a feasibility study. Environmental Health Perspectives (In Review) 1999.

XII. APPENDIX

Appendix A: Reclassification Scheme for CDWR/PUR comparison, Kings County, 1991

CDWR Crop-Code	CDWR Description	CDWR Area, Acres	Comparable PUR Code	PUR Description	PUR TotAc (Class)
4001	ALFALFA_&_ALFALFA_MIXTURES	57,974.2	23001	ALFALFA (FORAGE - FODDER) (ALFALFA HA'	41,783.9
6012	ALMONDS	2,974.8	3001	ALMOND	2,767.6
6001	APPLES	567.4	4001	APPLE	534.9
6002	APRICOTS	125.5	5001	APRICOT	276.5
5002	ASPARAGUS	959.7	16002	ASPARAGUS (SPEARS, FERNS, ETC.)	576.0
1001	BARLEY	31.8	29103	BARLEY, GENERAL	4,955.2
5003	BEANS_(GREEN)	345.2	15003	BEANS, SUCCULENT (OTHER THAN LIMA)	1,144.0
			28001	BEANS (ALL OR UNSPEC)	
3010	BEANS__DRY_(ALL_TYPES)	392.7	15001	BEANS, DRIED-TYPE	226.0
6003	CHERRIES	12.7	5002	CHERRY	75.6
3006	CORN	24,336.7	22005	CORN (FORAGE - FODDER)	9,564.9
			29119	CORN, HUMAN CONSUMPTION	
3001	COTTON	232,441.5	29121	COTTON, GENERAL	230,143.2
5016	FLOWERS__NERSERY_&_CHRISTMAS_TREE	569.2	151	N-GRNHS GRWN CUT FLWRS OR GREENS	217.4
			155	N-GRNHS GRWN TRNSPLNT/PRPGTV MTRL	
7009	JOJOBA	41.2	27018	JOJOBA (OIL CROP)	10.0
7008	KIWIS	281.6	6018	KIWI FRUIT	302.0
5008	LETTUCE_(ALL_TYPES)	202.7	13045	LETTUCE, HEAD (ALL OR UNSPEC)	320.0
5009	MELONS__SQUASH__AND_CUCUMBERS_(AI	2,489.3	10002	CANTALOUPE	2,161.0
			10011	PUMPKIN	
			29122	MELONS	
			10008	WATERMELONS	
6010	MISCELLANEOUS_DECIDUOUS	692.9	3008	PECAN	517.9
			6012	PERSIMMON	
			6015	POMEGRANATE (MISCELLANEOUS FRUIT)	
			4004	QUINCE	
5018	MISCELLANEOUS_TRUCK	646.8	11001	EGGPLANT (ORIENTAL EGGPLANT)	22.2
			11000	FRUITING VEGETABLES (ALL OR UNSPEC)	
			154	N-OUTDR CONTAINER/FLD GRWN PLANTS	
			29137	TURNIP, GENERAL	
7006	OLIVES	945.7	28014	OLIVE (ALL OR UNSPEC)	672.0
5010	ONIONS_AND_GARLIC	2,414.0	14007	GARLIC	1,720.2
			14011	ONION (DRY, SPANISH, WHITE, YELLOW, RED, ETC.)	
6005	PEACHES_AND_NECTARINES	6,774.0	5003	NECTARINE	6,276.3
			5004	PEACH	
6006	PEARS	104.8	4003	PEAR	34.9
6014	PISTACHIOS	6,082.8	3011	PISTACHIO (PISTACHE NUT)	4,550.2
6007	PLUMS	2,669.4	5005	PLUM (INCLUDES WILD PLUMS FOR HUMAN	2,367.2
3002	SAFFLOWER	47,619.8	29129	SAFFLOWER, GENERAL	40,198.6
3005	SUGAR_BEETS	5,731.4	29135	SUGARBEET, GENERAL	6,147.5
5015	TOMATOES	11,075.6	11005	TOMATO	13,764.3
			29136	TOMATOES, FOR PROCESSING/CANNING	
8000	Vineyards	4,770.2	29141	GRAPES	3,502.5
			29143	GRAPES, WINE	
6013	WALNUTS	6,982.2	3009	WALNUT (ENGLISH WALNUT, PERSIAN WAL	4,560.9

In PUR, not in CDWR

Crop-Code	Description	CDWR Area, Acres	Comparable PUR Code	PUR Description	
1002	Wheat	0.0	29139	WHEAT, GENERAL	13,801.2
1003	Oats	0.0	29125	OATS, GENERAL	941.5
3007	Grain Sorghum	0.0	29131	SORGHUM/MILO GENERAL	42.3
5006	Carrots	0.0	29111	CARROTS, GENERAL	242.0
5021	Peppers (Chilli, Bell, etc.)	0.0	11003	PEPPERS (FRUITING VEGETABLE), (BELL,CI	7.5
5022	Broccoli	0.0	13005	BROCCOLI	574.6
5023	Cabbage	0.0	13007	CABBAGE	5.0
5024	Cauliflower	0.0	13008	CAULIFLOWER	121.0
1006	Misc. and mixed grain and hay	0.0	22000	FORAGE - FODDER GRASSES (ALL OR UNSI	281.0
6008	Prunes	0.0	5006	PRUNE	26.0

In CDWR, not in PUR

Crop-Code	Description	CDWR Area, Acres	Comparable PUR Code	PUR Description	
1606	AIRPORT_RUNWAYS	798.4	67001	Airport	0.0
5007	CELERY	109.4	28003	Celery, General	0.0
2203	DAIRIES	3,993.1	43030,71005,	Dairies	0.0
6009	FIGS	16.3	6005	Figs	0.0
9000	Idle Lands	4.2	66002, 28108	Idle Lands	0.0
4003	MIXED_PASTURE	2,351.8	28035	Psture, General	0.0
4004	NATIVE_PASTURE	1,374.3	28035	Psture, General	0.0
7003	ORANGES	60.1	2006	Oranges, general	0.0
5011	PEAS	203.6	29127	Peas, General	0.0
2204	POULTRY_FARMS	857.4	55000	Poultry Farms	0.0
1603	RAILROAD	1,850.7	67005	Railroads	0.0
1200	Residential	21.6	68002	Urban/Residential	0.0
1100	Urban	14,142.2			
1411	STEEL_AND_ALUMINUM_MILLS	76.6	31418	Alumiunum Plant	0.0
3008	SUDAN	856.8	22011	Sudan	0.0
5013	SWEET_POTATOES	101.9	14018	Sweet Potatoes	0.0
4007	TURF_FARMS	254.0	33008	Torf, General	0.0
2300	Water	7,786.1	29673		
1402	EXTRACTIVE_INDUSTRIES	206.5	67009	INDUSTRIAL SITES (LUMBER YARDS, TANK I	0.0
1400	misc industrial	203.8			
1300	Commercial	35.0	77000	COMMERCIAL, INSTITUTIONAL OR INDUSTR	0.0
1400	misc industrial	203.8			
1305	INSTITUTIONS	220.6			
		35,728.0			

In PUR, No CDWR Match

Crop-Code	Description	CDWR Area, Acres	Comparable PUR Code	PUR Description	
			10	STRUCTURAL PEST CONTROL	1.0
			66000	UNCULTIVATED AGRICULTURAL AREAS (ALI	3,368.3
			67000	UNCULTIVATED NON-AG AREAS (ALL OR UN	209.0
					3,578.3

In CDWR, No corresponding PUR Code

Crop-Code	Description	CDWR Area, Acres	Comparable PUR Code	PUR Description
1504	CEMETERIES_-_IRRAGATED	119.5		
2201	FARMSTEADS	4,051.4		
2202	LIVESTOCK_FEED_LOTS	1,206.1		
1201	SINGLE_FAMILY_DWELLINGS_-_LOT_SIZE_1	1,194.9		
1601	UNPAVED_AREAS	1,242.1		
1600	Vacant	124.9		
3011	Misc_Field_Crops	43,177.8		
1006	Miscellaneous Grain	37,984.6		
1800	NATIVE_VEGETATION	279,443.5		
9001	Cropped_within_the_paSt_three_yearS	65,739.0		
9002	LandS_being_prepared_for_crop_production	26.1		
		434,309.9		