INTEGRATING LANDSAT AND CALIFORNIA PESTICIDE EXPOSURE ESTIMATION AT AGGREGATED ANALYSIS SCALES: ACCURACY ASSESSMENT OF RURALITY

by

Trang Minh VoPham

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

May 2014

DEDICATION

This thesis is dedicated to my family - especially Matthew David Weaver - for their relentless support, and to my University of Pittsburgh epidemiology doctoral mentor, Joel L. Weissfeld, MD, MPH, who engrained in me a great love and appreciation for all things geospatial.

ACKNOWLEDGEMENTS

I sincerely thank my thesis advisor, John Wilson, PhD, and committee members, Darren Ruddell, PhD and Tarek Rashed, PhD, for their invaluable expertise and guidance.

TABLE OF CONTENTS

Appendices

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Active ingredient
CA	California
CDPR	California Department of Pesticide Regulation
CDWR	California Department of Water Resources
CIR	Color-infrared
CO-MTRS	County, meridian, township, range, and section
CSL	Crop signature library
DDE	dichlorodiphenyldichloroethylene
DDT	dichlorodiphenyltrichloroethane
DEM	Digital Elevation Model
EPA	Environmental Protection Agency
ETM+	Enhanced Thematic Mapper Plus
FIPS	Federal Information Processing Standard
FSA	Farm Service Agency
GCP	Ground control point
GIS	Geographic information system
GloVis	Global Visualization
GPS	Global Positioning System
ISODATA	Iterative Self-Organizing Data Analysis Technique
L1T	Level 1 Standard Terrain Correction product
LPGS	Level 1 Product Generation System
MAUP	modifiable areal unit problem

MODIS	Moderate Resolution Imaging Spectroradiometer
MS	Multispectral
MSS	Multispectral Scanner
NAD83	North American Datum (1983)
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared band
NLAPS	National Land Archive Production System
OCP	Organochlorine
OP	Organophosphate
PCA	Principal component analysis
PEG	Parkinson's Environment and Gene study
PLSS	Public Land Survey System
PUR	Pesticide Use Report
R	Red band
RGB	red-green-blue
RHRC	Rural Health Research Center
RUCA	Rural-Urban Commuting Area
SSD	Sum of squared difference
SRS	Stratified random sampling
TM	Thematic Mapper
UA	Urbanized Area
UC	Urban Cluster

USGS	U.S. Geological Survey
ZCTA	ZIP Code Tabulation Area
ZIP code	Zone Improvement Plan code
WGS84	World Geodetic System (1984)

ABSTRACT

Pesticide exposure estimation in epidemiologic studies can be constrained to analysis scales commonly available for cancer data - census tracts and ZIP codes. Research goals included (1) demonstrating the feasibility of modifying an existing geographic information system (GIS) pesticide exposure method using California Pesticide Use Reports (PURs) and land use surveys to incorporate Landsat remote sensing and to accommodate aggregated analysis scales, and (2) assessing the accuracy of two rurality metrics (quality of geographic area being rural), Rural-Urban Commuting Area (RUCA) codes and the U.S. Census Bureau urban-rural system, as surrogates for pesticide exposure when compared to the GIS gold standard. Segments, derived from 1985 Landsat NDVI images, were classified using a crop signature library (CSL) created from 1990 Landsat NDVI images via a sum of squared differences (SSD) measure. Organochlorine, organophosphate, and carbamate Kern County PUR applications (1974-1990) were matched to crop fields using a modified three-tier approach. Annual pesticide application rates (lb/ac), and sensitivity and specificity of each rurality metric were calculated. The CSL (75 land use classes) classified 19,752 segments [median SSD 0.06 NDVI]. Of the 148,671 PUR records included in the analysis, Landsat contributed 3,750 (2.5%) additional tier matches. ZIP Code Tabulation Area (ZCTA) rates ranged between 0 and 1.36 lb/ac and census tract rates between 0 and 1.57 lb/ac. Rurality was a mediocre pesticide exposure surrogate; higher rates were observed among urban areal units. ZCTA-level RUCA codes offered greater specificity (39.1-60%) and sensitivity (25-42.9%). The U.S. Census Bureau metric offered greater specificity (92.9-97.5%) at the census tract level; sensitivity was low (≤6%). The feasibility of incorporating Landsat into a modified three-tier GIS approach was demonstrated. Rurality accuracy is affected by rurality metric, areal aggregation, pesticide chemical class, and pesticide exposure cutoff. Future research should explore integrating Landsat for higher spatial resolution pesticide exposure estimation.

CHAPTER ONE: INTRODUCTION

The goals of this research were: (1) to demonstrate the feasibility of using Landsat remote sensing in improving the spatiotemporal resolution of an established GIS-based pesticide exposure method modified to accommodate cancer data analysis scales; and (2) to evaluate the validity of two different measures of rurality as indicators of ZIP Code Tabulation Area (ZCTA)- and census tract-level pesticide exposure and determine which measure of rurality is a more accurate surrogate of pesticide exposure.

The feasibility of integrating Landsat remote sensing and pesticide exposure estimation is defined as demonstrating that agricultural pesticide application data can be matched to Landsat-derived crop fields. Specifically, an established GIS-based method matches California pesticide application data to land use survey agricultural crop fields to subsequently estimate human pesticide exposure (Rull and Ritz 2003). As land use surveys are intermittently updated for California counties every seven to 10 years, Landsat remote sensing, which captures Earth imagery every 16 to 18 days, provides an opportunity to produce agricultural crop fields during years lacking land use surveys (USGS 2013b). Demonstrating that pesticide application data - not matching to land use survey crops fields - can be matched to Landsat imagery classified into agricultural crops bolsters the feasibility of using remote sensing in pesticide exposure estimation.

An accuracy assessment of rurality addresses potential exposure misclassification when using rurality to quantify pesticide exposure – which is important to consider in the context of epidemiologic studies. Epidemiologic studies seeking to elucidate the relationship between pesticide exposure and human health outcomes have employed rurality as a surrogate measure of pesticide exposure, which takes advantage of the high prevalence of agricultural pesticide applications occurring in rural geographic areas (Franklin and Worgan 2005). Usage of rurality will inevitably misclassify some geographic areas as pesticide-exposed and vice versa, and the exact quantification of such misclassification has never before been addressed.

Location is the fundamental focus of the spatial sciences - a multifaceted force that influences human society (Waller and Gotway 2004). One avenue through which location impacts society is through the environment playing a direct role in human health. Where an individual lives and where an individual works has a direct effect on future health outcomes (Pickle et al. 2005). The environment is defined as exogenous factors nonessential to the normal functioning of human beings, and includes physical, chemical, and biological agents, in addition to the social, cultural, and political factors interacting with these agents (Rothman et al. 2008). Especially for individuals living in rural areas, the environment is associated with potential residential exposure to chemicals (Ward et al. 2000; Franklin and Worgan 2005). One group of chemicals that has played a prominent role in adversely affecting human health is pesticides (Alavanja et al. 2004). Pesticides are chemicals used to treat pests, such as insects (EPA 2012). In both 2006 and 2007, approximately 5.1 billion lb of total pesticides were used in the U.S. (EPA 2011). Pesticides have been specifically linked to the development of many types of cancers (Alavanja et al. 2004), which pose a large public health burden as the second leading cause of death in the U.S. (CDC 2012). In 2012, there were an estimated 1,638,910 cases of cancer diagnosed in the U.S. (NCI 2012). An approach that combines methods and concepts from epidemiology and the spatial sciences will be able to better understand the exact role of pesticides in cancer development and to confront the significant public health burden of cancer.

1.1 Harnessing Epidemiology and the Spatial Sciences

Epidemiology is a branch of science that seeks to elucidate the role that an exposure, such as pesticides, may play in the development of disease (Szklo and Nieto 2007). Epidemiology can involve carrying out research studies designed to provide an unbiased measure of association between a purported exposure and risk of a particular disease. Results from epidemiologic studies serve many segments of society, from informing policy makers to being a platform upon which other researchers build their research. The capacity to provide an unbiased measure of how an exposure is *truly* associated with a disease partly hinges on the validity of the measure used to indicate the exposure. Validity is the extent to which a measure is an indicator of what it is intended to measure (Szklo and Nieto 2007). Compared to a gold standard, the validity of a surrogate measure can be quantified through determining its sensitivity and specificity. Sensitivity refers to the capacity of a measure (e.g. rurality) to correctly identify features with a characteristic of interest [e.g. ZCTA- or census tract-level agricultural application of pesticides] (Szklo and Nieto 2007). Specificity refers to the capacity of a measure to correctly identify features without a characteristic of interest. A surrogate measure with low sensitivity will produce more false negatives, while a surrogate measure with low specificity will produce more false positives. Depending on how frequently the exposure occurs in the general population and whether sensitivity and/or specificity is affected, an epidemiologic study using a particular surrogate measure may report an exposure conferring less risk for an outcome than it truly does, or greater risk (i.e. biasing results towards or away from the null hypothesis of no association) (Szklo and Nieto 2007).

However, research endeavors are often limited in resources and time, and surrogate measures are frequently employed. Specifically, the main limitation of determining human exposure to pesticides has been inadequate methods of ascertaining past exposure (Franklin and Worgan 2005). Methods of determining pesticide exposure are either qualitative (e.g. self-reported pesticide exposure, occupation, etc.) or quantitative (e.g. direct biological measurement). A frequently used surrogate measure of pesticide exposure has been rurality, defined as the quality of a geographic area being rural (Alavanja et al. 2004; Rural Assistance Center 2012). Rurality may be self-reported (i.e. qualitative) or determined using various objective criteria, such as population density (i.e. quantitative). Rural areas are typically associated with agricultural activities, which are a primary source of pesticide exposure (Franklin and Worgan 2005). Pesticides are applied on agricultural lands and residential proximity to applications may contribute to pesticide exposure by way of applied pesticides drifting from intended sites (Ward et al. 2000; Rull and Ritz 2003). However, rurality is an imperfect surrogate of pesticide exposure likely misclassifying some rural areas as pesticide-exposed when they are truly not and vice versa.

On the other hand, GIS-based pesticide exposure methods are quantitative approaches that are increasing in usage (Ritz and Rull 2008; Maxwell et al. 2010a), which offer an objective alternative to qualitative measures, capable of using data from multiple databases containing relevant pesticide information and determining historical exposure through incorporation of many years of data (Alavanja et al. 2004). GIS-based methods represent a potentially cost-effective approach to ascertaining pesticide exposure compared to other methods, such as measurement and determination of serum levels of a pesticide. A GIS-based approach is also superior to qualitative measures, such as self-reported pesticide exposure, which are subject to recall bias (Alavanja et al. 2004; Franklin and Worgan 2005). GIS-based pesticide exposure methods have expanded to incorporate remote sensing, such as satellite-borne imagery, due to the spatiotemporal resolution offered by these data types (Maxwell et al. 2010b; Maxwell 2011). However, there exists a paucity of literature using remotely sensed data in cancer research examining environmental exposures, which stands to gain from the high spatiotemporal resolution of remote sensing data in reconstructing historical exposures (Maxwell et al. 2010a).

Pesticides pose potentially harmful human health effects and have been associated with the development of chronic diseases, such as cancer (Dich et al. 1997; Alavanja et al. 2004). In order to adequately determine whether or not pesticide exposure is *truly* associated with a health outcome in epidemiologic studies, whether or not a surrogate measure used to indicate pesticide exposure is *truly* valid must first be addressed. GIS, guided by fundamental principles underlying epidemiology and the spatial sciences, provides a powerful way to combine relevant spatial and non-spatial data to address this research question in ways that would be not otherwise possible. This research addresses two gaps in the literature regarding: (1) the lack of cancer epidemiology studies using remote sensing in environmental exposure assessment observed in Maxwell, Meliker, et al. (2010); and (2) the absence of research examining the validity of rurality as an indicator of pesticide exposure.

CHAPTER TWO: BACKGROUND

Pesticides are pervasive chemicals designed to be toxic to organisms, such as insects, herbs, and fungi (Blair 1988; EPA 2012), and are grouped into functional classes according to which organisms they control, such as insecticides, herbicides, and fungicides. Pesticides are also further categorized into chemical classes according to their chemical structure and biological mechanisms of action, such as organochlorine pesticides (OCPs) (e.g. DDT and endosulfan) (Alavanja et al. 2004). Pesticides are composed of their active ingredient, in addition to other ingredients, such as solvents, which comprise the pesticide products available in the market. Pesticides are used most frequently in agriculture, horticulture, and vector control (e.g. antimalarial), followed by forestry and livestock production (Dich et al. 1997). Exposure to pesticides occurs through direct, higher-level routes, such as through occupation, and indirect, lower-level but more frequent routes, such as through drinking water, food, air, and dust. In the U.S., the primary source of exposure is from consumption of dairy, fish, meat, and poultry products (Ritz and Costello 2006). Food is potentially contaminated with pesticides during production, storage, and/or transport processes (Oates and Cohen 2011).

2.1 Residential Proximity to Agricultural Pesticide Applications

An important source of pesticide exposure occurs through residential proximity to pesticide applications on agricultural lands, as pesticides may drift from intended sites through the ground and the air via spray drift and post-application volatilization to far locations (i.e. pesticide residues change from liquid to gas/vapor form in various climatic conditions) (Rull and Ritz 2003; Alavanja et al. 2007; Ritz and Rull 2008; EPA 2009).

The most vulnerable populations are those residing in rural areas and farming families (Ward et al. 2000), who are potentially exposed through dermal contact and ingestion of pesticides in household dust and in groundwater (Gunier et al. 2001). Farm families frequently reside within 100 yd (91 m) of crop fields (Ward et al. 2000). Aerial pesticide applications can drift between 500 and 1,000 m and boom-type sprayers [application via spray nozzles at regular intervals (Ministry of Agriculture 2013)] can drift between 300 and 800 m (Ward et al. 2000). Pesticides can enter homes through drift and dust on shoes and farmers' clothing (Ritz and Rull 2008). Residential proximity, as a route of exposure, poses a potentially large threat to human health, as pesticides are less likely to degrade and volatilize in homes due to the absence of moisture, sunlight, and microorganisms, and are able to persist over time (Ritz and Rull 2008; Gunier et al. 2011).

Although residential proximity to agricultural pesticide applications is itself a surrogate of actual human exposure, it has been directly tied to levels of pesticides in carpet dust samples (Gunier et al. 2011). Carpet dust samples from 89 residences in agricultural areas in California were collected in a determinants of exposure study (Franklin and Worgan 2005). Gunier et al. (2011) demonstrated that residential proximity to agricultural pesticides, measured as the application of six pesticides within a 1,250 m residential buffer [geographic coordinates captured using Global Positioning System (GPS) device] using pesticide application and land use data 730 days before dust collection was significantly correlated ($p<0.05$) with concentrations of pesticides in carpet dust (ng/g). Spearman rank correlation coefficients ranged between 0.23 and 0.50. Therefore, measurement of residential proximity to agricultural pesticide applications, captured using a GIS incorporating spatial information, is a potentially meaningful way to capture human exposure.

2.2 Evolution of Pesticide Exposure Estimation

Substantial progress has been achieved in pesticide exposure assessment by moving from self-reported measures towards more sophisticated, objective metrics. However, the majority of epidemiologic studies investigating the association between pesticide exposure and human health have employed interview-administered methods, which are associated with various limitations - most prominently recall bias (Franklin and Worgan 2005). Most exposure metrics have also relied on surrogates of true human exposure (Nuckols et al. 2004). In the context of pesticides, environmental concentration is the presence of the pesticide in a carrier medium, such as in the air. Exposure concentration is the presence of the pesticide at the point of contact, such as in the zone of breathing. Dose is the amount of the pesticide that enters the human body (i.e. absorbed). In epidemiologic studies, measures of environmental concentration have typically served as surrogates for exposure concentration and dose.

Pesticide exposure metrics are categorized as qualitative or quantitative (Franklin and Worgan 2005). Qualitative metrics are derived from questionnaires and interviews, such as self-reported occupational history, residential locations, and exposure to pesticides (e.g. garden and residential use). Specific examples include occupation as a farmer (ever and duration), type of crop raised (duration and acres), and application of any pesticide (ever and duration) (Alavanja et al. 2004). However, a limitation of qualitative metrics is the absence of identification of specific pesticides. Recall also varies according to subpopulations, such as farmers likely having better recall as they may directly participate in the purchase and/or application of pesticides. In comparison, migrant farmworkers and those occupying pesticide-treated residences may not be able to recall, or may be unaware of, specific pesticide names (Alavanja et al. 2004). To address the limitation of recall bias, self-reported measures can be supplemented with a review from experts, such as occupational hygienists.

Quantitative metrics are derived through direct measurement of external exposure from the air (i.e. environmental monitoring), or from biological markers in serum, urine, fat, etc. (i.e. biological monitoring/biomonitoring) (Franklin and Worgan 2005; Alavanja et al. 2007). Measuring pesticide levels in carpet dust is an example of environmental monitoring. Biomonitoring is considered the gold standard approach that provides a measure of human pesticide exposure from all pathways and routes, and is advantageous when the chemical of interest has a long biological half-life and when its concentration is not affected by disease. Quantitative metrics also include usage of exposure databases, such as those collecting information for pesticide regulation purposes. Integrated pesticide exposure metrics have also been developed, which combine self-reported information with other relevant data, such as personal protective equipment, to better estimate exposure (Alavanja et al. 2004). For example, job-exposure matrices (JEMs) are integrated metrics that are typically region-based (e.g. British JEM) and incorporate information regarding job title, tasks, and industry to estimate exposure intensity. Exposure intensity algorithms are an extension of JEMs, which weight cumulative pesticide exposure by chemical- and applicator-specific information, such as work practices.

2.2.1 GIS-Based Pesticide Exposure Metrics

Built upon the concern of high and persistent pesticide exposure among rural residents, inadequacies of frequently employed qualitative metrics, and the potentially harmful effect of pesticides on human health, there has been a burgeoning body of research that focuses specifically on utilizing the concepts and techniques of the spatial sciences to improve pesticide exposure assessment. Geographic information system (GIS)-based approaches represent a quantitative method of pesticide exposure ascertainment integrating different sources of spatial and non-spatial information, such as the California Department of Water Resources (CDWR) land use surveys and the California Department of Pesticide Regulation (CDPR) Pesticide Use Reports (PUR) database (Alavanja et al. 2004; Nuckols et al. 2004; Franklin and Worgan 2005). GIS-based metrics improve upon the limitations of existing methodologies. Specifically, recall bias, prominent in qualitative methods, is minimized through combining objectively acquired information, such as remotely sensed data. GIS-based metrics can be used to determine pesticide exposure levels for the general population, as individuals are likely unaware of agricultural pesticides close to their residence (Ward et al. 2000; Alavanja et al. 2007). Using GIS also represents a cost-effective and time-efficient approach to assessing exposure, compared to collecting and measuring biological samples in a large enough study sample with adequate statistical power to detect meaningful differences. Furthermore, many pesticides have short biological half-lives, and their biologic measurement, though useful in assessing recent exposure, may be irrelevant in attempting to determine past exposure that may have precipitated a chronic disease (Franklin and Worgan 2005).

2.2.2 Advantages of GIS-Based Metrics in Cancer Epidemiology

GIS-based pesticide exposure metrics are especially powerful tools in the context of cancer epidemiology. Methods of measuring pesticide exposure should consider many important underlying issues in relation to the study of chronic diseases, such as cancer (Franklin and Worgan 2005). Cancer is frequently associated with long latency periods (i.e. time between first exposure and clinical diagnosis of disease), typically 20 years or more (Blair 1988; Rothman et al. 2008). Historical reconstruction of past exposure is important in capturing the potential effect of a latency period. When possible, exposure assessment should precede the onset of disease to showcase a temporal relationship, which is important evidence of a causal relationship (Alavanja et al. 2004; Szklo and Nieto 2007). Multiple routes of exposure exist, such as dermal, inhalational, and oral. Furthermore, individuals are potentially exposed to a variety of pesticides. Specifically, agricultural workers are likely exposed to multiple pesticides over the crop-growing season. Depending on the crop type, pesticides can be applied in combination via tank mixes and over multiple time points during the growing season, which pose difficulties in determining the impact of a specific pesticide on disease (Franklin and Worgan 2005).

GIS-based pesticide exposure metrics are able to address all of the aforementioned issues. Through incorporating multiple data sources with locational information and specific chemicals, often spanning long time periods, meaningful and relevant measures of human exposure to pesticides can be derived. The following are potential sources of pesticide exposure that are addressed in using a GIS-based metric: inhalation of ambient air, persistence in household dust, “take-home” of pesticides from occupations, soil drift, groundwater contamination, dermal contact in fields, and direct ingestion of contaminated produce (Gunier et al. 2001). Most, if not all, of these potential sources of exposure are associated with residential proximity to agricultural pesticide applications.

2.2.3 Improvements in GIS-Based Pesticide Exposure Methods over Time

GIS-based pesticide exposure metrics have improved over time through addressing fundamental concepts underlying the spatial sciences. GIS allows for the capacity to combine many spatial data sources, which are often associated with different data representations (i.e. data models), scales, and levels of accuracy (Nuckols et al. 2004). Spatial data can be represented as vector data models, which represent entities as points, lines, and polygons, and are typically associated with real-world phenomena with discrete, unambiguous boundaries, and as raster data models, such as satellite imagery, representing data through pixels, or cells, which are better-suited for continuous phenomena (Waller and Gotway 2004). Data sources are available at different scales, or spatial resolutions/granularities (i.e. smallest distinguishable and/or mappable unit). Scale can also refer to analysis scale (i.e. how phenomena are measured/aggregated) and operational/phenomenon scale (i.e. scale at which a process of interest operates) (Montello 2001; Nuckols et al. 2004). A data source may be aggregated to a particular analysis scale that is not relevant to the underlying geographic process of interest it is attempting to represent. Lastly, error can emanate from positional error (i.e. inaccuracies in locational information), attribute error (i.e. inaccuracies in data describing specific locations), and temporal error (i.e. mismatches in temporal currency of data).

Geospatial pesticide exposure methods have taken these fundamental issues into account. Different data models necessitate different types of analytic tools, and advances in technology and technical knowledge have facilitated the use and analysis of different data types in GIS environments. Multiple data sources, often times collected for purposes unrelated to research, have been combined to ultimately provide improved spatiotemporal resolution. Specifically, the field of GIS-based pesticide exposure assessment has grown to include high-resolution remote sensing technology, such as aerial photographs and Landsat satellite imagery, land use surveys, and pesticide exposure databases, to improve the spatiotemporal resolution of capturing individual-level residential exposure to agricultural pesticides.

2.2.4 Inception of GIS-Based Pesticide Exposure Metrics: Crop Maps

Pioneering the use of GIS in pesticide exposure assessment,Ward et al. (2000) conducted a feasibility study to determine the extent to which Landsat satellite imagery could be used to reconstruct historical crop patterns. Validated using Nebraska Farm Service Agency (FSA) aerial photographs with annotated crop information, a historical land cover map of Adams, Buffalo, and Hall Counties in Nebraska was created using a Landsat multispectral (MS) image from 1984. Six agricultural land cover types (i.e. corn, sorghum, soybeans, alfalfa, rangeland, and bare soil) were screen-digitized using the FSA records. Crop-specific probabilities of pesticide use were determined using information from surveyed farmers of the University of Nebraska Agricultural Extension Service and usual number of applications of each pesticide from the U.S. Environmental Protection Agency (EPA) Biological and Economic Analysis Division of the Office of Pesticide

Programs. After creating 500 m buffers around geocoded [i.e. assigning a geographic location to an address record (Waller and Gotway 2004)] residences of study subjects from a non-Hodgkin lymphoma study, exposure to crop pesticides applied to one or more major crop types and the (average) distance from each residence to crop field centroid(s) within the buffer were calculated.

Ward et al. (2000) demonstrated that Landsat remote sensing could provide useful information relevant to studies seeking to quantify pesticide exposure. For example, the authors showed rural residences (*N*=10; outside of town boundaries) had a greater number of crop fields proximate to their homes and were closer in distance to crop fields compared to community residences (*N*=97; located within a town boundary). Specifically, 100% of rural residences vs. 15% of community residences had at least one crop field within a 500 m residential buffer. The median distance to crop field centroids was 378.3 m for rural residences vs. 419.9 m for community residences.

Ward et al. (2006) extended the previous work to determine if there is an association between residential proximity to agricultural fields and indoor pesticide concentrations that could adversely affect human health. The authors evaluated the utility of crop maps to predict crop herbicide levels from residential carpet dust samples. Using collected vacuum cleaner dust from study subjects of a non-Hodgkin lymphoma study in Iowa, 14 herbicides were measured. Landsat MS images, validated using FSA records, were used to create land cover maps to identify corn and soybean fields between 1998 and 2000. Among 112 residences with locations recorded using GPS devices, 58% had crops within 500 m of their home. Sixty-one percent of rural residences had detectable levels of herbicides in carpet dust, compared to 15% of in-town residences. The odds of detecting at least one agricultural herbicide was 7.4 [95% confidence interval (*CI*) 1.3-41.3] times greater among residences with more than 300 ac of corn and soybean fields within 750 m compared to no crops within 750 m, adjusted for agricultural jobs.

There was also a significant increase in concentration of agricultural herbicides in carpet dust (ng/g) per ac increase of crops within 500 to 750 m of a residence [β (regression coefficient) 1.01; 95% *CI* 1.00-1.02], adjusted for other buffer distances. The geometric mean of agricultural herbicides measured in house dust increased with recent agricultural work, with levels of approximately 366 ng/g [geometric standard deviation (*SD*) 4.6] in homes with current agricultural workers, 121.9 ng/g (geometric SD 2.5) with former agricultural workers, and 111.5 ng/g (geometric *SD* 2.5) with no agricultural workers. The authors interpreted their findings as confirmation of the "take-home" pathway of exposure for families living with an agricultural worker who potentially exposes family members to pesticides from clothing, shoes, etc.

2.2.5 Utilizing the California Pesticide Use Reports (PURs)

Bell et al. (2001) and Gunier et al. (2001) forged more direct approaches to estimating potential residential pesticide exposure by utilizing the comprehensive California Pesticide Use Reports (PUR) database (CDPR 2013). Rather than assuming all crop fields are treated similarly with respect to pesticide applications, these authors shifted their focus towards California, which is both agriculturally productive and has legally required protocols for growers and applicators to report use of all restricted-use pesticides since 1974 and all pesticides since 1990 (Bell et al. 2001; Gunier et al. 2001; CDPR 2013). PURs include information regarding specific applications of pesticides, such as the name and pounds of pesticide active ingredient applied, acres treated, date of application, and Public Land Survey System (PLSS) section location of application. The PLSS system is used to divide and describe U.S. lands for surveying purposes, and imposes a grid of square sections measuring 1 mi on a side spanning the entire U.S. (National Atlas 2013).

Bell et al. (2001) examined the relationship between maternal residential proximity to pesticide applications and fetal death due to congenital anomalies across Fresno, Kern, Kings, Madero, Merced, Monterey, Riverside, San Joaquin, Stanislaus, and Tulare Counties in California. Restricted-use pesticides from five pesticide chemical classes between 1983 and 1984 were assessed. Using PUR-derived pesticide application locations at the PLSS section-level and maternal addresses located using county maps, broad and narrow geographic definitions of pesticide exposure were calculated. According to the broad definition, a mother was exposed to pesticides if a PUR application was within her section of residence, or the eight adjacent sections. A narrow definition only considered the section of residence. Despite the advantages of investigating the effects of specific pesticides and pesticide chemical classes, the authors did not geocode the exact locations of maternal residences. PLSS sections are 1 mi^2 in resolution, and distances between residences and pesticide applications could not be determined.

Gunier et al. (2001) demonstrated an improved usage of PUR data through calculating average annual pesticide application rates (lb of pesticide active ingredient per mi^2) between 1991 and 1994 for each PLSS section. The authors examined pesticide use based on groupings related to chemical classes and toxicological evidence (i.e.

probable carcinogens, possible carcinogens, genotoxic compounds, and reproductive or developmental toxicants) for all census block groups across California. Pesticide use at the PLSS section level was allocated to census block groups based on section area within each census block group. Census block group-level pesticide use density was calculated after dividing by the census block group area. Gunier et al. (2001) also developed a methodology to weight annual pesticide application rates according to the pesticide's potential to cause cancer (i.e. carcinogenicity) using U.S. EPA classifications and exposure potential via volatilization and environmental persistence. Using this GIS-based approach, the authors found that most census block groups in California (57-99%) averaged less than 1 lb per mi^2 of average annual use for each pesticide group and individual pesticide evaluated.

Several studies have since adopted the approach detailed in Gunier et al. (2001). Reynolds et al. (2002) conducted an ecologic study of childhood cancer cases between 1988 and 1994 in California in relation to census block group-level pesticide exposure. Reynolds et al. (2005) subsequently improved pesticide exposure assessment by incorporating a residential buffer around each of the geocoded addresses. Investigating the relationship between maternal residential proximity to pesticide applications and early childhood cancer, Reynolds et al. (2005) estimated pesticide exposure within a half-mile of geocoded maternal residences. Pounds of pesticide use were assigned to each study subject's residence based on the percentage area of each PLSS section within each half-mile buffer. Pounds were summed across the relevant time period of interest and divided by the buffer area (approximately 0.79 mi^2) for each pesticide toxicological group, chemical class, and individual pesticide. Despite the utility of using specific PUR pesticide application data in relation to a geocoded residential buffer, methods of improving the spatial resolution of the PUR data reported beyond the 1 mi^2 PLSS section level were needed. The authors noted that usage of PLSS section-level data is more sensitive in capturing potential pesticide applications, but potentially at the cost of specificity.

2.2.6 Enhancement of PUR-Derived Metrics Using Land Use Surveys

The next advance in GIS-based pesticide exposure methodologies using PUR data attempted to increase the specificity (i.e. minimization of false positives) of estimates through finding a relevant buffer distance around geocoded residences to capture pesticide drift, and through increasing the spatiotemporal resolution of determining which agricultural lands were applied with pesticides. Rull and Ritz (2003) laid the foundation for usage of the CDWR land use surveys to make use of the PUR attributes regarding crop type and field acreage associated with pesticide applications. County-based CDWR land use surveys are conducted every seven to 10 years to describe land use and crop cover, with a minimum mapping unit of 0.81 ha (0.003 mi^2) (Nuckols et al. 2007). Rather than basing estimates on residence within a 1 mi^2 PLSS section, resolution was improved to 1:24,000 or 1 in to 2,000 ft by incorporating land use surveys (Rull and Ritz 2003). Specifically, the authors determined the likely locations of crop fields (using CDWR land use surveys) near residences (within 500 or 1,000 m) upon which PUR pesticide applications took place. The authors accounted for the seasonal rotation of crops (i.e. same fields used for different crops) by collapsing seasonal field crops (e.g. cotton, grains, potatoes, tomatoes, and alfalfa) into a single field crop class. Rull and Ritz (2003) created a three-tier approach to assign pounds of applied pesticides to agricultural lands based on the certainty of a PUR crop type matching land use survey data (Goldberg et al. 2007). Annual application rates (lb/ac) were calculated by summing the applied pounds of pesticide (from PUR) divided by treated crop acres (from land use survey) intersecting a 500 or 1,000 m residential buffer.

Rull and Ritz (2003) compared their approach to the Bell et al. (2001) broad vs. narrow approach by generating 1,000 randomly selected samples of 200 addresses from residential parcel centroids in Kern County, California. The Rull and Ritz (2003) PUR and land use survey approach was designated as the gold standard, and a residence was considered exposed if a pesticide-treated field was within a 500 or 1,000 m buffer. The Bell et al. (2001) approach (broad: exposed if PUR application reported within section of residence or adjacent sections; narrow: exposed if PUR application reported within section of residence) and a land use survey-only approach (residential proximity to crop fields; exposed if crops grown within 500 m of residence) was compared to the gold standard. The authors demonstrated that measures of association between pesticide exposure and a health outcome would be attenuated, or biased towards the null hypothesis, if using lower resolution metrics [i.e. Bell et al. (2001) approach and land use survey-only approach] that did not combine both PUR and land use survey data. For various pesticides when compared to the gold standard, the Bell et al. (2001) broad definition was associated with perfect sensitivity (100%) but poor to good specificity (62-93.9%). Specificity was improved with the Bell et al. (2001) narrow definition (98.7-99.4%), though sensitivity decreased (35.3-54.8%). The land use survey-only model was associated with decreasing specificity and increasing sensitivity with increasing buffer size (500 m buffer: 60.1% sensitivity and 94% specificity; 1,000 m buffer: 72.2% sensitivity and 87% specificity).

Rull and Ritz (2003) also touched on the impact of residential mobility. Attenuation of measures of association becomes more pronounced with increasing exposure prevalence and increasing mobility rate. This issue is particularly problematic if individuals move to urban areas, which would decrease specificity and increase the number of false positives. Taken as a whole, Rull and Ritz (2003) demonstrated that usage of a higher resolution metric increases specificity and decreases the extent to which measures of association are attenuated, particularly when the true exposure prevalence is low in the population. Many epidemiologic studies have since adopted this approach for studying cancers and Parkinson's disease (Marusek et al. 2006; Rull et al. 2006a; Rull et al. 2006b; Roberts et al. 2007; Costello et al. 2009; Gatto et al. 2009; Ritz et al. 2009; Rull et al. 2009; Manthripragada et al. 2010; Cockburn et al. 2011; Wang et al. 2011; Lee et al. 2012, 2013). In practice, these studies have weighted pesticide application rates (lb/ac) by the proportion of the area of pesticide-treated acres intersecting a 500 m residential buffer.

The accuracy of the Rull and Ritz (2003) GIS-based approach of estimating individual-level residential exposure to agricultural pesticides was demonstrated in a validation study among participants of the Parkinson's Environment and Gene (PEG) study in Central California (Fresno, Kern, and Tulare Counties). For 22 Parkinson's disease cases and 22 age- and gender-matched Medicare controls and randomly selected residential parcels, exposure was defined as the weighted average of organochlorine pesticide applications (lb/ac) within a 1,000 m residential buffer between 1974 and 1999.

Using lipid-adjusted dichlorodiphenyldichloroethylene (DDE) levels measured in serum as the gold standard, the GIS-based metric was associated with 87% specificity and 38% sensitivity. The GIS-based metric, body mass index, age, gender, mixing and loading pesticides by hand (derived from occupational questionnaire data), and residential pesticide use explained 47% of the variance in DDE serum levels.

Building off of the Rull and Ritz (2003) approach, Nuckols et al. (2007) also utilized PUR and land use survey data to derive pesticide exposure, but at the crop level by not collapsing field crops. Collapsing crops may introduce additional issues in interpretation when the collapsed category does not include the crop type in the PUR database (Rull and Ritz 2003). For each PLSS section intersecting a 500 m residential buffer, the authors calculated an annual crop-specific pesticide application rate (lb/ac) between 1988 and 1994 by dividing the total amount of pesticides applied to each crop of interest during the time period by the total area of the crop field within the PLSS section. This was weighted according to the crop area within a 500 m buffer, and then divided by the area of the buffer to obtain a pesticide use density measure in lb per mi^2. Six pesticides were evaluated and residences of participants from a California Department of Health Services childhood cancer study were geocoded. Using this metric as the gold standard, the performance of a PUR-only metric (pesticide application rate weighted by the area of the PLSS sections within a 500 m buffer) was assessed. A residence was considered exposed if there was any pesticide use on a crop field within 500 m of a residence (gold standard), or pesticide use in a section within 500 m (PUR-only model). The authors also used an additional exposure cutoff of greater than the 25^{th} percentile of pesticide use. By using a coarser-scale, PUR-only metric, sensitivity is 100%. Nuckols et al. (2007) demonstrated good specificity with various pesticides [e.g. 96% specificity with dicofol (5% prevalence) compared to 86% specificity with propargite (15% prevalence)] using a cutoff of any pesticide use within 500 m. However, specificity decreased to between 29 and 45% when excluding residences with no pesticide use within 500 m and using a 25th percentile cutoff of exposure. The results of overall agreement mirrored that of specificity, where overall agreement between the gold standard and PUR-only metric was high (88-98%), but decreased when excluding residences with no pesticide use within 500 m (35-58%).

2.2.7 Downscaling PUR-Derived Metrics Using Landsat Satellite Imagery

Approaches to ascertaining pesticide exposure using PUR data have been further refined with remotely sensed data, which are data captured from a distance, such as aerial photographs and satellite imagery (Waller and Gotway 2004). Despite being a rich resource for large-scale data, few cancer epidemiologic studies have utilized remotely sensed data for environmental exposure assessment (Maxwell et al. 2010a). Remote sensing is particularly relevant to cancer epidemiology because it can be used to reconstruct environmental exposures and characterize environmental change. Landsat satellite imagery, for example, provides moderate to high resolution data spanning 39 years (USGS 2013b). One primary advantage is the multispectral and multitemporal features of Landsat data, which allow for landscape features to be distinguished based on spectral and phenological (i.e. seasonal changes in vegetation) characteristics (USGS 2011).

Maxwell et al. (2010b) demonstrated the potential use of Landsat imagery in improving pesticide exposure assessment in California. The authors showed how the spatiotemporal uncertainty regarding crop field-level changes due to the infrequent CDWR land use surveys could be addressed by using temporally varying Landsat imagery. Specifically, CDWR land use surveys represent a snapshot of the agricultural landscape at one point in time. During the interim time between land use surveys, if there is more than one survey, the shapes, sizes, and existence of crop fields may have changed. Furthermore, during a given crop-growing season, fields may be used for more than one crop (i.e. multi-cropped) and only a portion of a field in the CDWR land use survey may be utilized for growing a particular crop.

A time series of 24 Landsat 5 and 7 images in 2000 was collected for Fresno County, California. The time series of Landsat imagery was intersected with CDWR land use surveys to determine crop field locations. Normalized Difference Vegetation Index (NDVI) values, which measure vegetative growth, were derived from the Landsat images for 17 crop types. Maxwell et al. (2010b) demonstrated that variation in NDVI values across the year can showcase evidence of multi-cropped fields and potential misclassification of fields from the land use surveys (e.g. absence of NDVI-based evidence for vegetation within a land use survey-labeled crop field). Maxwell (2011) subsequently presented a case study of the exact methods used to execute a Landsat imagery-based approach to downscale, or improve the spatial resolution of, PUR data reported at the PLSS section level. PUR applications of the pesticide paraquat were selected for Fresno County, California, in 1994. Crop field boundaries based on similar phenological characteristics (i.e. NDVI values) were derived from Landsat imagery across different dates throughout 1994. Crop types were determined by comparing pixels from these delineated crop field boundaries to the crop signature library established in Maxwell et al. (2010b). It was shown that Landsat data could be used to identify PUR errors (e.g. PUR data indicates pesticide applications on non-existent crop fields) and to determine which exact area of a field was used for a particular crop type associated with a PUR pesticide application.

2.3 Surrogate Measures of Pesticide Exposure: Rurality

Despite the burgeoning research into geospatial methods of pesticide exposure ascertainment, data limitations, lack of technical knowledge, etc., may necessitate the use of crude surrogates, or proxy indicators, of pesticide exposure. Rurality, or the extent to which a geographic area is rural, has been used as an indicator of pesticide exposure (Alavanja et al. 2004). The rationale for using rurality stems from agricultural lands associated with pesticide applications being more common in rural areas (Ward et al. 2000; Alavanja et al. 2004; Franklin and Worgan 2005). Although some measures of rurality fall under the umbrella of qualitative pesticide exposure metrics (e.g. self-reported residence in a rural area), some are quantitative metrics that offer an objective alternative, such as through incorporating existing information regarding population density. In the context of epidemiology and investigating the association between pesticide exposure and one or more health outcomes, there are three primary issues regarding the use of rurality to indicate pesticide exposure that should be considered: (1) potential misclassification of pesticide exposure, (2) different definitions of rurality, and (3) variation in analysis scales.

2.3.1 Misclassification of Pesticide Exposure

Validity, or accuracy, is the extent to which a measure is an indicator of what it intends to measure (Szklo and Nieto 2007). Validity and exposure misclassification are related concepts, where the impact of validity is manifest in exposure misclassification, or information bias. Specifically, the validity of a pesticide exposure metric not only influences the accuracy of the metric in truly indicating pesticide exposure, but may inflate or obscure exposure-disease relationships (Franklin and Worgan 2005). These relationships can be quantified as measures of association in epidemiologic studies, such as odds ratios [i.e. odds of disease among individuals exposed to the purported exposure of interest compared to those not exposed (Szklo and Nieto 2007)]. The exact effect of using an inaccurate exposure metric on measures of associations depends on the extent of exposure misclassification and the prevalence of the exposure in the study population of interest.

Exposure misclassification can be understood in terms of the classic error model, where an exposure metric is measured with error and is an imperfect surrogate for the true exposure (Nuckols et al. 2004). The degree of misclassification can be measured using sensitivity (i.e. capacity of a measure to correctly identify features with a characteristic of interest) and specificity (i.e. capacity of a measure to correctly identify features without a characteristic of interest) (Szklo and Nieto 2007). For example, in the context of a comparative, epidemiologic study with two study groups - cancer cases and non-cancer controls - evaluating the association between pesticide exposure and cancer, nondifferential misclassification (i.e. extent of exposure misclassification does not differ between the study groups) of pesticide exposure will bias the measure of association towards the null hypothesis of no association. Furthermore, if the prevalence of pesticide exposure is low in a study population (<10%), then decreases in specificity will substantially attenuate the measure of association. However, for more frequent exposures, reductions in sensitivity is associated with greater bias (Szklo and Nieto 2007). Rurality likely misclassifies some geographic areas as exposed to pesticides when they are truly not and vice versa; however, the exact extent to which rurality may misclassify pesticide-exposed geographic areas remains unknown.

2.3.2 Variation in Rurality Definitions

Different definitions of how to define rurality exist, and different exposure-disease relationships may be observed depending on which definition is used (Rural Assistance Center 2012). For example, a rurality metric may only consider population information in delineating a rural geographic area, or may consider multiple factors, such as population and work commuting information. Although an analysis may compare and contrast results using different definitions, this approach may not meaningfully contribute to determining which rurality definition most adequately reflects pesticide exposure, or the processes underlying pesticide exposure.

2.3.3 Variation in Analysis Scales

Furthermore, variation in analysis scales of rurality may also influence study results. Analysis scales, or how data is measured/aggregated, may vary due to compulsory aggregation of cancer data, general data availability, ecologic study designs, and incorporation of contextual information in studying individual-level phenomena

(Montello 2001). Cancer data, such as that derived from cancer registries, are often aggregated to areal units (e.g. census tracts and ZIP codes) for the purposes of patient confidentiality (Boscoe et al. 2004; Waller and Gotway 2004). To avoid scale-translation issues, exposure data can be aggregated to identical areal units, such as to evaluate the potential association between census tract-level rurality (exposure) and census tract-level cancer incidence rates (outcome) (Boscoe et al. 2004). These aggregations form the fundamental units of analysis in the majority of studies employing geospatial techniques (i.e. ecologic studies) (Nuckols et al. 2004), where the unit of analysis is not the individual, but an aggregated unit (Szklo and Nieto 2007). Ecologic studies can be important in generating hypotheses; however, depending on how the data are aggregated, different results can be observed [i.e. modifiable areal unit problem (MAUP)] (O'Sullivan and Unwin 2010). Studies have also examined individual-level phenomena while incorporating both individual-level and ecologic, contextual variables to perform multi-level analyses (Jacquez 2004). This approach is meaningful in attempting to capture the effect of a variable that may operate at a scale beyond the individual. Irrespective of a study's unit of analysis as the individual or an ecologic aggregate, the underlying issue is that usage of different rurality metrics, whether based on different rurality definitions or aggregated to different scales, may lead to different results.

Taken together, rurality is an intuitive surrogate measure of pesticide exposure in that applications of agricultural pesticides frequently occur in rural geographic areas. However, analytic results may vary according to usage of different rurality definitions and investigations at different analysis scales. Most importantly, a rurality metric is inevitably associated with inaccuracies, likely misclassifying some geographic areas as pesticide-exposed that are truly not and vice versa. In many study populations, such as the state of California, the overall frequency of pesticide exposure is low [2.2% of California population reside in rural areas (USDA 2013b)]. In the context of researching the relatively infrequent exposure of residential proximity to agricultural applications of pesticides, the impact an exposure metric with suboptimal specificity (i.e. high number of false positives) will bias the results of a study towards the null hypothesis and attenuate the true exposure-disease relationship. Therefore, determining the validity/accuracy of a rurality-based exposure metric as an indicator of pesticide exposure, as well as understanding the impact of using different rurality definitions at varying analysis scales, is important in elucidating its performance and adequacy as a surrogate measure of true pesticide exposure.

CHAPTER THREE: METHODS AND DATA SOURCES

In order to focus the analysis, the study area of interest was Kern County, California and historical pesticide exposure between 1974 and 1990 was calculated. The study area and time period of interest were constrained by the California Department of Pesticide Regulation (CDPR) Pesticide Use Report (PUR) database, as pesticide reporting to the CDPR began in 1974 (CDPR 2000b). As there are a wide array of pesticides in use throughout California, the analysis focused on three pesticide chemical classes previously associated with primary liver cancer - organochlorine pesticides (OCPs), organophosphates (OPs), and carbamates (Cordier et al. 1993; Ezzat et al. 2005; Persson et al. 2012).

For each census tract and ZCTA in Kern County, annual pesticide chemical class-specific application rates (lb/ac) were calculated using a GIS. Pounds of applied pesticides were derived from the PUR database and crop field acreage from land use surveys, classified Landsat imagery, and PLSS sections. A new GIS-based pesticide exposure methodology is presented, which modifies the Rull and Ritz (2003) three-tier approach combining PURs, CDWR land use surveys, and PLSS sections to estimate census tract- and ZCTA-level pesticide exposure and to incorporate Landsat imagery. A crop signature library (CSL) of Normalized Difference Vegetation Index (NDVI) values was created using Landsat imagery in 1990, which was used to classify segments derived from Landsat NDVI images in 1985 into agricultural crop fields.

Rurality was measured for each ZCTA and census tract using two common metrics, Rural-Urban Commuting Area (RUCA) codes and the U.S. Census Bureau urban-rural classification system. A statistical analysis, including calculating measures of validity (i.e. sensitivity and specificity), was performed to formally evaluate the extent to which ZCTA- and census tract-level rurality metrics are valid indicators of pesticide exposure, as well as to determine which surrogate measure offers greater accuracy. All GIS-related geoprocessing and visualization was performed in ArcGIS 10.1 and IDRISI Selva; statistical analyses was performed in SAS 9.3 .

3.1 Research Hypotheses

Rural designations using RUCA codes and the U.S. Census Bureau urban-rural classification system were hypothesized to be less sensitive and less specific compared to the GIS-based pesticide exposure metric (i.e. gold standard) in assigning ZCTA- and census tract-level pesticide exposure. The RUCA code system was hypothesized to be a more accurate surrogate measure of pesticide exposure compared to the U.S. Census Bureau urban-rural classification system due to its incorporation of both population and work commuting information. The U.S. Census Bureau urban-rural classification system only incorporates population information. In other words, RUCA codes, by virtue of their definition, were hypothesized to better reflect areas truly rural where agricultural applications of pesticides are more likely to occur.

3.2 Study Area: Kern County, California

California is the third largest state in the U.S., 158,706 mi^2 in size with 58 counties (CA.gov 2013). The most populous cities are Los Angeles, San Diego, San Jose, San Francisco, and Fresno. In 2012, there were 38,041,430 individuals residing in California, 2.2% (836,441 individuals) of whom were rural residents (USDA 2013b). In 2007, over

25% of the total statewide land area was devoted to farmland. California is the most agriculturally productive state in the U.S. in terms of farm output and productivity (Economic Research Service 2012).

In 2007, approximately 3% of all California farms were located in Kern County (*N*=2,117) (USDA 2007a). Kern County is one of 19 counties nestled in California's agriculturally intensive Central Valley, which produces 25% of the food Americans consume (Figure 1) (NPR 2002). Over 2 million ac of Kern County were devoted to farmland, with an average size of 1,116 ac per farm. Over 68% of Kern County farms were used for cropland (*N*=1,449; 942,827 c), of which 81% (*N*=1,169; 764,929 ac) were devoted to harvested cropland and 15% (*N*=222; 41,081 ac) to pasture grazing (USDA 2007c). A total of 836 farms were used as orchards (407,208 ac).

Figure 1 Kern County, California, study area of interest
(Data from U.S. Census Bureau 2013)

In 2011, 191 million lb of pesticide active ingredients were used in California (CDPR 2011a). Over 28 million lb were used in Kern County, ranking it as the second highest county in the state for pesticide usage after Fresno County. A total of 945 Kern County farms reported using chemicals to control insects, followed by 735 for weed/grass/brush control, 120 for nematode control, and 471 for disease in crops and orchards (USDA 2007b). The most frequently pesticide-treated agricultural commodities in 2011 included grapes, carrots, oranges, and pistachios (Table 1) (CDPR 2011b).

Table 1 Common pesticide-treated crops in Kern County, 2011[1]

Crop	**Rank**	**Applied pesticides (lb)**	**Pesticide applications (*N*)**	**Treated land (ac)**
Almonds	1	7,996,450	31,468	2,990,073
Grapes	2	4,227,137	40,740	2,096,170
Carrots	3	3,170,438	4,203	243,140
Oranges	4	2,476,753	12,092	639,149
Pistachios	5	2,032,358	7,248	699,843

[1] Data adapted from CDPR (2011b)

3.3 Data Sources

3.3.1 Pesticide Exposure Data

Table 2 lists the data sources that were used to execute the research methodology. The CDPR PUR database is the most comprehensive pesticide reporting system in the world, collecting data regarding agricultural pesticide use throughout California (CDPR 2013). Between 1974 and 1989, commercial pest control operators (e.g. structural applicators) were required to report all pesticide use and farmers were required to report restricted pesticide use, or pesticides with high potential to cause public health harm. Since 1990, a full-use reporting system has been adopted. The PUR database focuses on agricultural pesticide applications, but also includes applications to parks, golf courses, etc. PUR information includes the name and pounds of pesticide active ingredient applied, field and crop acreage treated, date of application, and PLSS section of application. The PUR database contains 45,000 pesticide products; 1,000 new products are added each year and 1,000 are inactivated due to nonrenewal, suspension, or cancellation.

PUR data are reported at the PLSS section level (Rull and Ritz 2003). The PLSS system was introduced earlier and divides the country into townships measuring six miles on a side and these, in turn, are subdivided into 36 1 mi^2 sections (National Atlas 2013). PLSS surveys start at an initial point from which townships are surveyed north, south, east, and west. The north-to-south line running through the initial survey point is called the principal meridian for that PLSS survey. The east-to-west line running through the initial point is called the baseline, and is perpendicular to the principal meridian. Townships are identified through a township designation (i.e. north or south of baseline) and range designation (i.e. east or west of principal meridian). In California, PLSS sections are uniquely identified according to their county, principal meridian, township, range, and section. There are 8,455 PLSS sections intersecting Kern County (Figure 2).

The CDWR conducts land use surveys of agricultural lands for California counties focusing on over 70 crop types. Aerial photographs, satellite imagery, and GPS devices are used to delineate crop field boundaries. County-based surveys are conducted every seven to 10 years (CDWR 2013). The earliest land use survey conducted in Kern County occurred in 1990 (Figure 3).

Table 2 Data sources[1]

Dataset	Description	Geographic extent	Data type	Spatial resolution	Temporal currency
CDPR PURs	Agricultural pesticide use database	California, U.S.	Text files	Reported at PLSS (1 mi^2) section level	1974 to present
PLSS sections	Cadastral dataset	California	Vector data model (polygon)	PLSS polylines defined by survey points accurate to ≥40 ac level	Updated in 2011
CDWR land use surveys	Surveys of agricultural lands	California	Vector data model (polygon)	Minimum mapping unit of 0.003 mi^2	1976 to present; updated every 7-10 years
RHRC RUCA codes Version 2.0	Census tract and ZIP code rural/urban designations	U.S.	Text files	Reported at ZCTA and census tract level	2000
U.S. Census Bureau TIGER/Line shapefiles	Administrative delineations: census tracts and ZCTAs	U.S.	Vector data model (polygon)	200 ft resolution	2000
U.S. Census urban-rural classification	Urbanized Areas (UAs) and Urban Cluster (UCs)	U.S.	Vector data model (polygon)	200 ft resolution	2000
USGS and NASA Landsat imagery	Remotely sensed satellite imagery	Global	Raster data model	30 m for red and near infrared spectral bands (Landsat 4 and 5 Thematic Mapper sensors)	1972 to present

[1] Data from RHRC (2000); U.S. Census Bureau (2000, 2013); Cal-Atlas Geospatial Clearinghouse (2013); CDPR (2013); CDWR (2013), National Atlas (2013); and USGS (2013b)

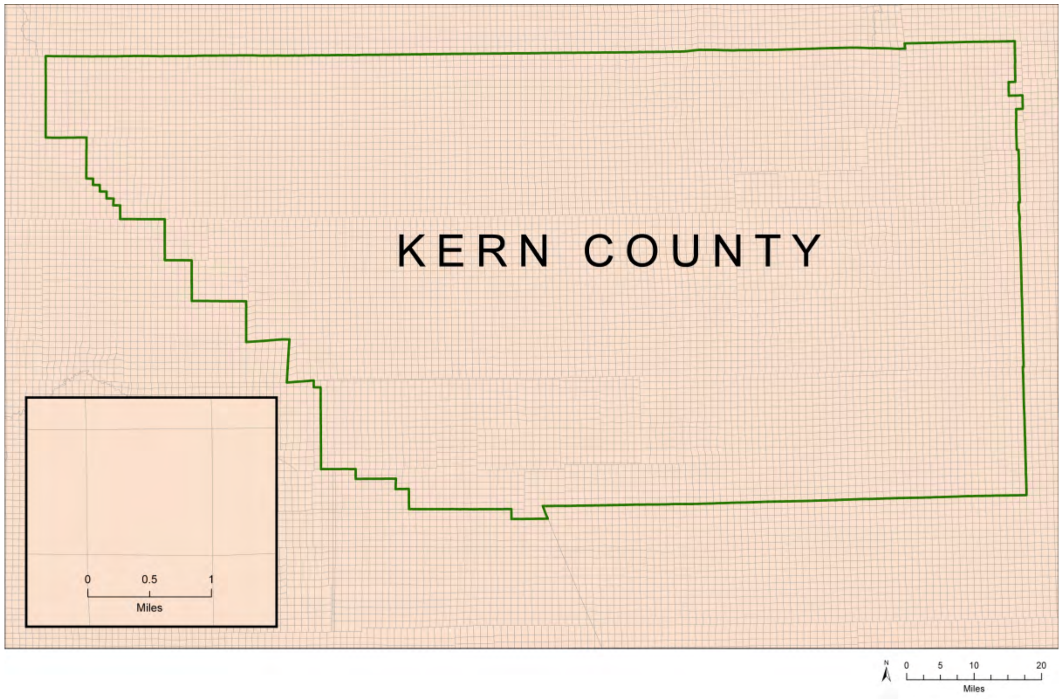


Figure 2 PLSS sections in Kern County
(Data from Cal-Atlas Geospatial Clearinghouse 2013; and U.S. Census Bureau 2013)

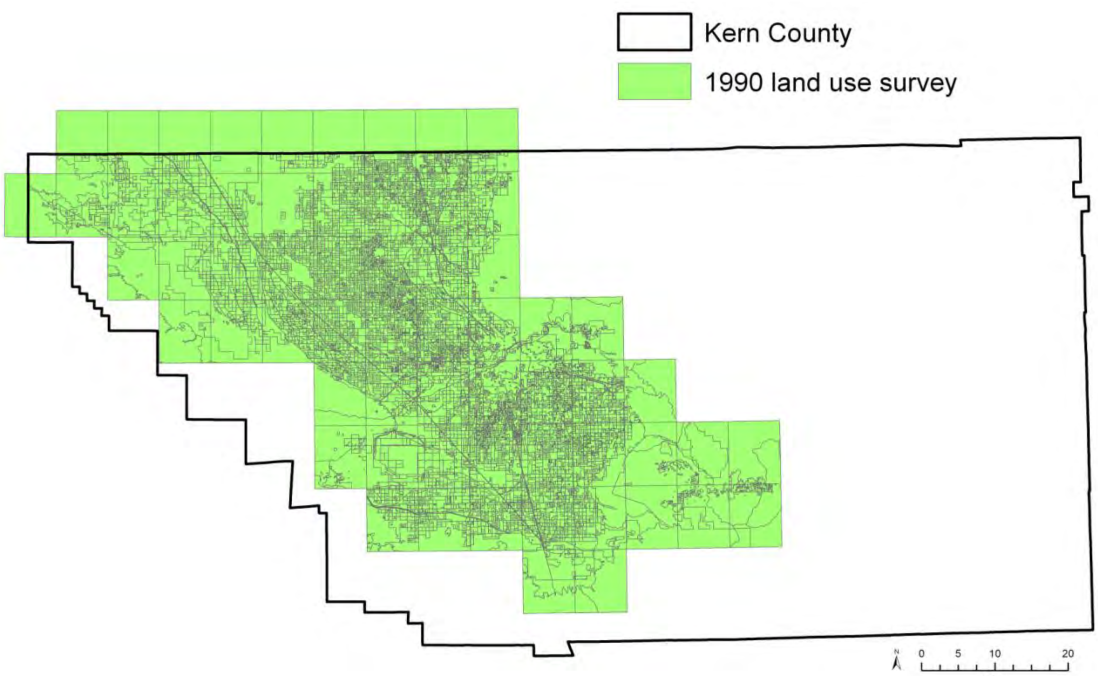


Figure 3 Map showing geographic extent of 1990 Kern County land use survey
(Data from CDWR 2013; and U.S. Census Bureau 2013)

3.3.2 Landsat Imagery

The Landsat program was started by the U.S. Geological Survey (USGS) and the National Aeronautics and Space Administration (NASA) to continuously collect Earth imagery (USGS 2013b). The first Landsat satellite, Landsat 1 or the Earth Resources Technology Satellite (ERTS), was launched in 1972. The Return Beam Vidicon (RBV) and Multispectral Scanner (MSS) sensors were onboard Landsat 1 through 3. Landsat 4 and 5 launched in 1982 and 1984, respectively, included the Thematic Mapper (TM) and MSS sensors (Table 3). Landsat 7 included the Enhanced Thematic Mapper Plus (ETM+) sensor, and Landsat 8, launched in 2013, carries the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) sensors (Maxwell et al. 2010a). The Landsat 4 and 5 remote sensing system characteristics, relevant to this analysis, are listed in Table 4.

Table 3 Landsat 4 and 5: Thematic Mapper (TM) sensor[1]

Band number	Spectral range (μm)	Wavelength	Spatial resolution (m)
1	0.45 - 0.52	Blue-green	30
2	0.52 - 0.60	Green	30
3	0.63 - 0.69	Red	30
4	0.76 - 0.90	Near infrared	30
5	1.55 - 1.75	Mid infrared	30
6	10.4 - 12.5	Far infrared	120
7	2.08 - 2.35	Mid infrared	30

[1] Data adapted from Campbell and Wynne (2011)

The NDVI is the most commonly used vegetation index to characterize vegetative growth, or greenness (i.e. relative vegetative density and health) (USGS 2011). NDVI values can be derived from Landsat satellite imagery using the red (R) and near infrared

Table 4 Landsat 4 and 5 remote sensing characteristics[1]

Characteristic	Description
Inception	Landsat 4: 1982 Landsat 5: 1984
Revisit frequency	16 days
Orbit	Near-polar, sun-synchronous
Swath width	185 km
Geographic extent	Global
Sensors	Multispectral Scanner (MSS), Thematic Mapper (TM)
Geographic reference	UTM coordinate system, WGS84 datum
Applications	Earth observation (EO)

1 Data adapted from USGS (2013b)

(NIR) bands: NIR-R/NIR+R. The NDVI system uses the wavelengths of light absorbed and reflected by green plants captured in Landsat satellite sensors; reflectance properties change as the growing season progresses. NDVI values range between -1.0 and +1.0. NDVI values less than 0.1 indicate areas with barren rock, sand, or snow (i.e. sparse vegetation). NDVI values between 0.2 and 0.6 indicate moderate vegetation, and values beyond 0.6 indicate dense vegetation.

3.3.3 Rural-Urban Commuting Area Codes

The Rural Health Research Center (RHRC) RUCA codes (Version 2.0) classify ZCTAs and census tracts using a 33-code scheme that incorporates 2000 U.S. Census Bureau daily work commuting, population density, and urbanization information. The RUCA system is comprised of a two-level classification system (RHRC 2000). The first level consists of whole numbers between 1 and 10 corresponding to metropolitan, micropolitan, small town, and rural area categories (USDA 2012). Second-level subgroups reflect secondary commuting flows associated with the first-level categories. Different RUCA code groupings exist to classify geographic units as urban or rural.

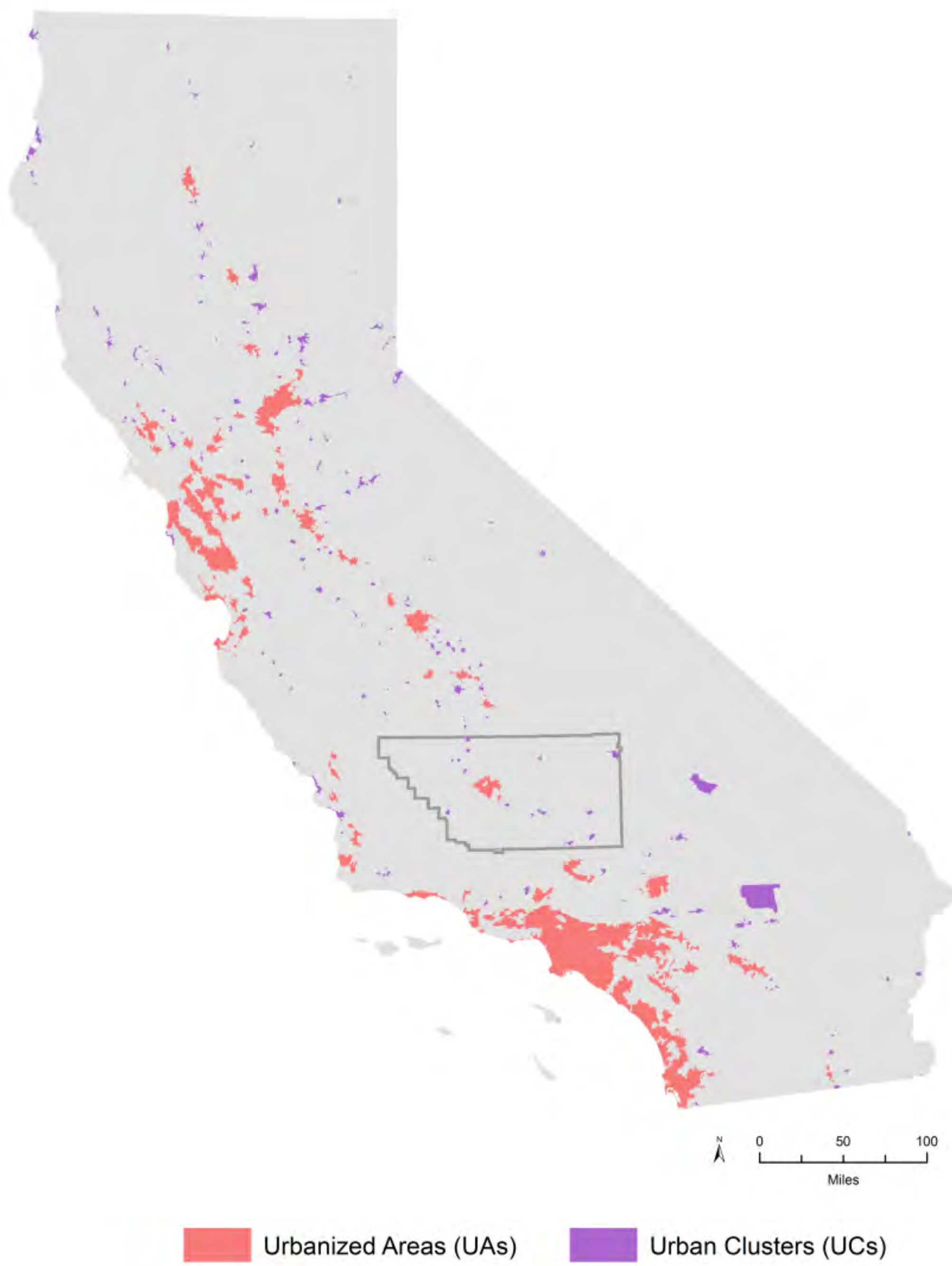


Figure 4 Urbanized Areas and Urban Clusters across California, 2000
(Data from U.S. Census Bureau 2013)

3.3.4 U.S. Census Bureau Urban-Rural Classification

The 2000 U.S. Census Bureau urban-rural classification system categorizes geographic areas across the U.S. as urban or rural. Specifically, Urbanized Areas (UAs) are geographic areas with a densely settled core of census block groups or census blocks with a population density of at least 1,000 individuals per mi^2, surrounding census blocks with a population density of at least 500 individuals per mi^2, and with a total population of 50,000 or more (Figures 4 and 5). Urban Clusters (UCs) are geographic areas with a densely settled core of census block groups or census blocks with a population density of at least 1,000 individuals per mi^2, surrounding census blocks with a population density of at least 500 individuals per mi^2, and with a total population of at least 2,500 individuals, but less than 50,000 (U.S. Census Bureau 2000). Any geographic areas outside of UAs and UCs are considered rural. Data from the year 2000 were chosen to compare with the RUCA code assignments.

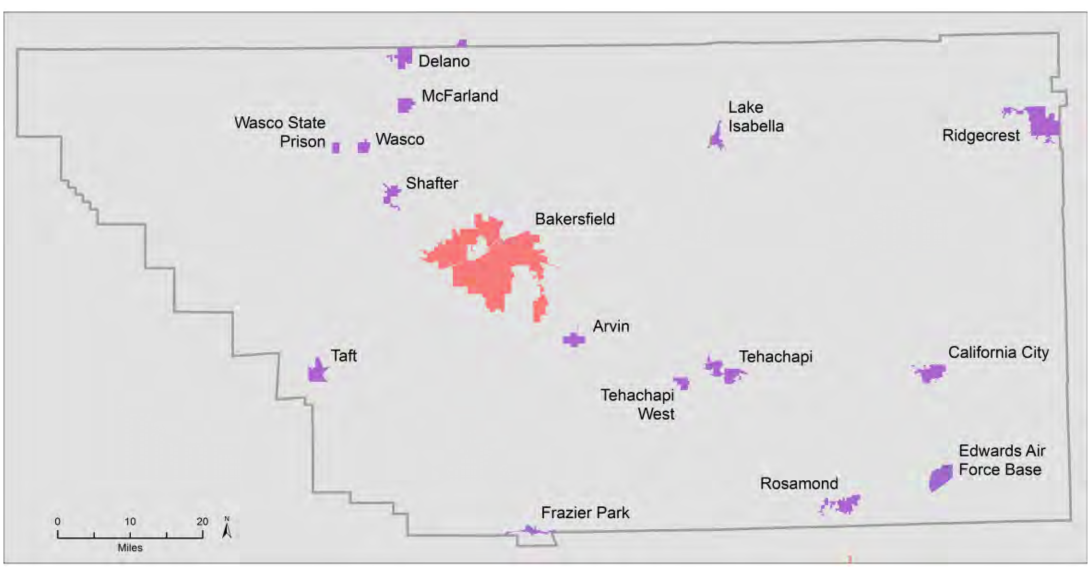


Figure 5 Kern County UAs and UCs, 2000 (Data from U.S. Census Bureau 2013)

3.4 Pesticide Exposure Estimation

3.4.1 Preparation of PUR, PLSS, and Land Use Survey Data

All PUR files between 1974 and 1990 were downloaded. Each PUR record contains information regarding an individual active ingredient used in a pesticide application. Since a pesticide product may contain multiple chemicals, there may be multiple PUR records for a single pesticide application (CDPR 2000b).

The workflow for processing PUR, land use survey, and PLSS data for use in the GIS environment is illustrated in Figure 6. Using agricultural pesticide references (Dich et al. 1997; Gunier et al. 2001; Alavanja et al. 2004; Greene and Pohanish 2005; Rull et al. 2006a, 2009; Wood 2010; AgroPages 2013), a database of pesticides belonging to the organochlorine (OCP), organophosphate (OP), and carbamate chemicals classes was compiled - including pesticide name and CDPR chemical code. Unique identifiers

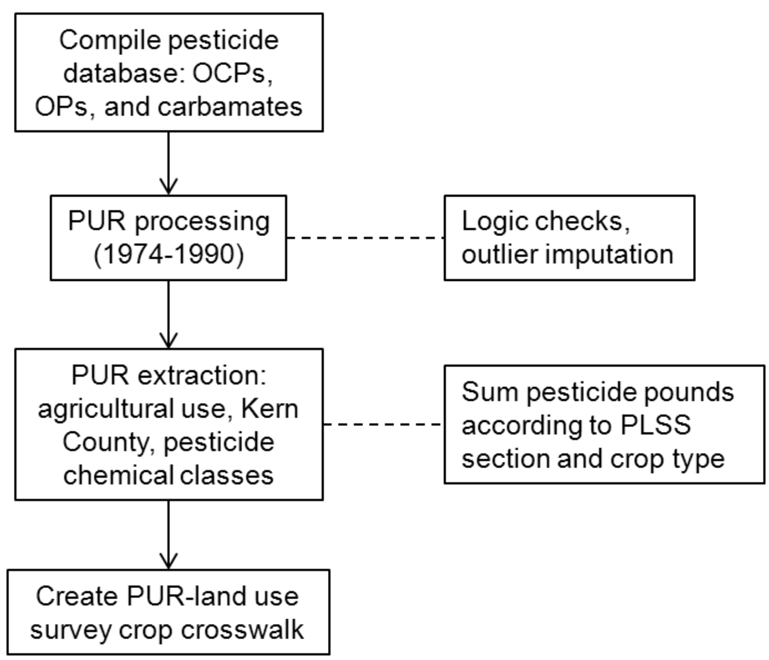


Figure 6 Methodological workflow: PUR, land use survey, and PLSS processing

comprised of the CDPR county code, principal meridian, township, range, and section (CO-MTRS) were created to combine PUR data with PLSS and land use survey data.

Per CDPR data quality standards (CDPR 2000a), the following PUR logic checks were performed separately for each year between 1974 and 1989 (Appendix B; Table B1): (1) duplicates and (2) spatially inconsistent county (using CO-MTRS) outside county boundary. The following logic checks were performed for 1990 PUR data: (1) duplicates, (2) spatially inconsistent county, (3) inconsistent county code, (4) missing agricultural field location identifiers, (5) inconsistent CO-MTRS for a location, (6) inconsistent acres planted, and (7) treated acres greater than planted acres. The CDWR land use survey logic check was omitted due incorporating land use survey information in the tiered PUR matching methodology. Not all CDPR logic checks were performed on PUR data between 1974 and 1989 because the logic checks were created for PUR data from 1990 onward. Therefore, some of the variables required for the logic checks, such as the grower identification number, were available starting in 1990. Depending on the logic check definition, PUR records identified using a logic check were either excluded from the analysis or the first record was retained.

To increase comparability between PUR and CDWR land use survey data, in addition to addressing the uncertainty regarding the seasonal rotation of crops (e.g. double-cropping) and intercropping, the following crop types in both datasets were collapsed into a single field crop category: (1) grain and hay crops; (2) field crops; (3) pasture; and (4) truck, nursery, and berry crops (Rull and Ritz 2003; Rull et al. 2006a). A crosswalk between PUR commodity codes (different codes were used between 1974 and 1989 compared to 1990 onward) and CDWR land use survey crop codes (1981-1992 and

1993-1997 legends) was created. Executive decisions regarding CDPR commodity codes assigned to CDWR land use survey crops that were not exact name matches were documented [e.g. CDPR tangerines (pre-1990 code 2104; 1990 code 2008) assigned to CDWR land use survey oranges (code C3)].

Between 1974 and 1989, outliers were identified as pesticide application rates meeting two of the three CDPR outlier flag definitions created for PUR data beginning in 1990 (CDPR 2002): (1) pesticide application rates [lb active ingredient (AI)/treated ac] greater than 200 lb/ac (greater than 1,000 lb/ac if fumigation) (only considers PUR records reported in acres), and (2) pesticide application rates (lb AI/treated unit) (considers all PUR records) greater than 50 times the median rate for all uses of that pesticide product [identified using manufacturing firm number, label sequence number, revision number, and registration firm number between 1974 and 1983, and as EPA registration number between 1984 and 1989], commodity code, unit type, and record type (production agriculture vs. monthly report). Outliers were identified in 1990 PUR records using three CDPR-created outlier flags (the two aforementioned outlier definitions in addition to neural network). All outliers were imputed with the statewide median rate for the pesticide AI in that year; pounds of AI were recalculated.

Between 1974 and 1989, agricultural, non-summary PUR records were extracted; in 1990, all daily and monthly production agriculture PUR records were extracted. The following selections were made from PUR records: those associated with an OCP, OP, or carbamate in the compiled pesticide database and applied in Kern County. As a result of the logic checks, only PUR records with a valid CO-MTRS identifier were included.

Pounds of applied AI were then summed according to crop type and CO-MTRS (Figure 6).

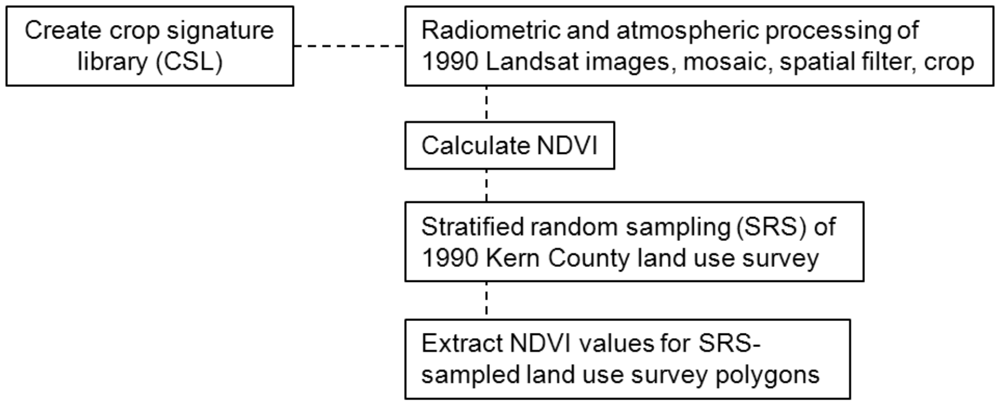


Figure 7 Methodological workflow: Landsat remote sensing and crop signature library

3.4.2 Incorporation of Landsat Imagery: Crop Signature Library (CSL)

A crop signature library (CSL) was compiled using a time series of Landsat 4 and Landsat 5 Thematic Mapper (TM) imagery acquired between January and October 1990 for Kern County - 1990 is the year in which the earliest Kern County CDWR land use survey is available as a ground truth (Figure 7) (Maxwell et al. 2010b; Maxwell 2011; CDPR 2013). Landsat images for November and December were not available in 1990. Images from Paths 41 and 42 and Rows 35 and 36 were requested from USGS Global Visualization (GloVis) Viewer as they cover the geographic extent of Kern County (Figure 8). Images with excessive cloud cover were excluded. All images were Standard Terrain Correction [Level 1T (L1T)] products, which are radiometrically and geometrically processed images using ground control points (GCPs) and a Digital Elevation Model (DEM) (USGS 2013a). All images were associated with a common coordinate system: Universal Transverse Mercator (UTM) Zone 11N (WGS84 datum; meter).

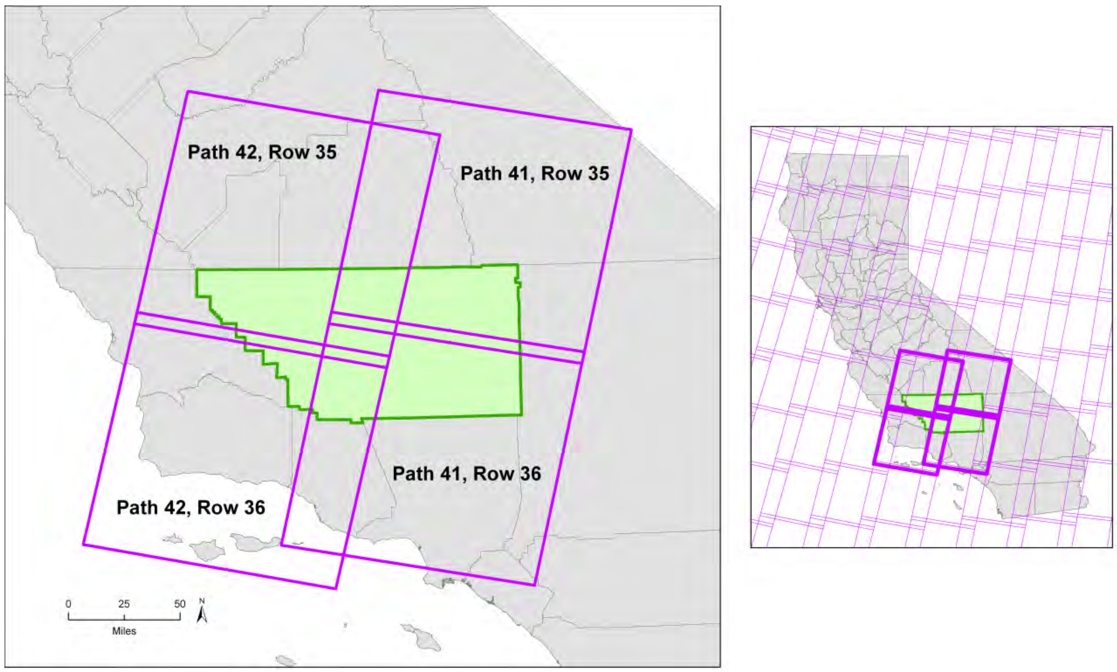


Figure 8 Landsat Path-Row scenes intersecting Kern County
(Data from U.S. Census Bureau 2013; and USGS 2013c)

Using IDRISI Selva, TM images for band 3 (red) and band 4 (near infrared) were radiometrically corrected to at-sensor reflectance using published radiometric calibration coefficients and image metadata [MTL files for Level 1 Product Generation System (LPGS)-processed images] (Chander et al. 2009). Atmospheric correction was executed using the Chavez cosine estimation of atmospheric transmittance (COST) model, taking into account date, time of day, band center wavelength, gain, bias, cosine of the solar zenith angle (90-solar elevation angle), and assuming the downwelling spectral irradiance, path radiance due to haze (i.e. digital number of objects with zero reflectance, such as deep clear lakes), and spectral diffuse sky irradiance is zero (Campbell and

Wynne 2011). Chavez (1996) developed the COST, or Cos(t), model to integrate the Dark Object Subtraction (DOS) model for haze removal, in addition to estimating absorption from Rayleigh scattering and atmospheric gases. For each month in 1990, all four Path/Row scenes (where available) were mosaicked one Path at a time (cover overlap method matching on grey level using non-background values). Clouds near the Path 41-to-42 overlapping region were masked out before a mosaic to join the two Paths. Subsequent to mosaicking, negative reflectance values, potentially associated with random error related to water and/or shadows, were recoded to a reflectance value of 0 (YCEO 2013). A median spatial filter (3x3 kernel) was applied to each mosaic to minimize random noise (Vassiliou et al. 1988; Mather and Koch 2011), and the mosaic was cropped to a smaller geographic area enclosing Kern County. NDVI values were calculated using bands 3 and 4. NDVI images were re-projected to the California Teale Albers (NAD83 datum; meter) coordinate system (30 m spatial resolution; nearest neighbor resampling to not alter pixels).

Guided by a natural color [red-green-blue (RGB) band combination] multispectral (MS) Landsat 5 image of California provided by Cal-Atlas (Cal-Atlas Geospatial Clearinghouse 2013), polygons representing the geographic extent of each monthly NDVI image not affected by clouds or shadows were digitized and intersected to create a cloud-free zone. Using the 1990 Kern County CDWR land use survey, stratified random sampling (SRS) eligibility criteria for land use survey polygons included (1) single-use (i.e. not double- or triple-cropped, intercropped, or mixed); (2) at least 4 ha in area (Maxwell et al. 2010b); and (3) intersecting the cloud-free zone. SRS using strata defined by land use classes (e.g. C1=grapefruit) was performed to select at most 30 eligible polygons per stratum. Non-agricultural land use classes, such as native vegetation (NV), were included in the CSL to facilitate the discrimination between as many land use classes as possible during subsequent classification.

Negative NDVI values, indicative of an absence of green vegetation, were recoded to 0 (Beck et al. 2006). Separately for each month between January and October 1990, NDVI values for each pixel were extracted using a mask defined by the SRS-sampled land use survey polygons. NDVI values for points straddling multiple crop field boundaries were deleted. The median NDVI value for each land use survey polygon was calculated; the median NDVI value for each land use class for each month was retained in the CSL for subsequent classification (Figure 9).

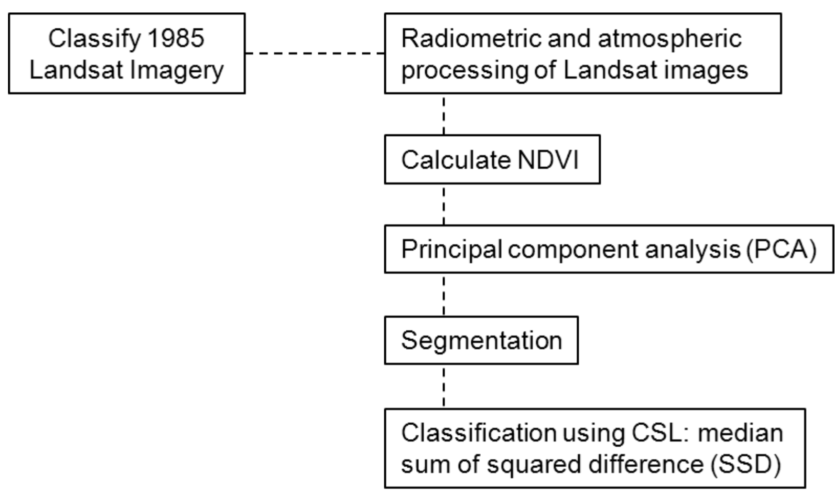


Figure 9 Methodological workflow: classification of Landsat images

3.4.3 Classification of 1985 Landsat Imagery

Using the CSL, NDVI Landsat images from 1985 for Paths 41 and 42 and Rows 35 and 36 were radiometrically and atmospherically processed and digitally enhanced to facilitate classification according to land use class using a sum of squared differences (SSD) measure (Figure 9). An earlier time period of 1985 was chosen to address the PUR records occurring before 1990. All images were acquired using the Landsat 5 TM sensor, processed with either the LPGS system (MTL metadata) or the National Land Archive Production System (NLAPS) (WO metadata) system (Chander et al. 2009) and associated with the UTM Zone 11N coordinate system (WGS84 datum; meter). Landsat images between January and October 1985 were requested from GloVis to parallel the CSL. Due to the absence of images for Path 42 (majority of Kern County agricultural fields in this Path) in February 1985, this month was excluded from classification.

Images were corrected to at-sensor reflectance (Chander et al. 2009), cloud-masked, mosaicked, smoothed using a median filter (3x3 kernel), and cropped to a geographic area enclosing Kern County. NDVI values were derived using the red and near infrared bands 3 and 4, respectively. Cloud- and shadow-free geographic areas within the NDVI images available for all months (except February) in 1985 were digitized and intersected to create a segmentation-eligible zone.

A principal component analysis (PCA) was performed on the nine monthly NDVI images as a data compression method (Maxwell 2011; Lippitt et al. 2012). Principal components 1, 2, and 3 (Maxwell 2011) were used as inputs for an object-based segmentation to delineate crop field boundaries - and other land features in the selected geographic extent. Segmentation groups adjacent pixels into segments according to spectral homogeneity (Campbell and Wynne 2011). Segmentation was performed iteratively using different input parameters, comparing resultant segmentation products to crop field boundaries according to a color composite of the first three principal components (PC1: red, PC2: green, PC3: blue) and an August 1985 Landsat image displayed using a color-infrared (CIR) band combination (USDA 2013a). CIR imagery is useful when examining crop field boundaries and irrigated vegetation (CDOC 2013). The following parameters were used for the final segmentation output: window of 3, tolerance of 80, weight mean factor of 0.5, and weight variance factor of 0.5. All datasets were re-projected to the California Teale Albers coordinate system (NAD83 datum; meter).

Using all NDVI pixel values for each segment, the resultant segmentation vector shapefile was classified using the 1990 CSL according to a distance measure: the smallest sum of squared differences (Maxwell 2011) using the median NDVI value for each segment compared to the median value of all CSL land use classes for each available month in 1985 (January, March to October). A sensitivity analysis was performed comparing resultant classified segments when using a CSL with (1) all land use classes; (2) land use classes except broad groupings (e.g. F for field crop, no subclass given); or (3) land use classes except broad groupings and SRS land use strata with less than 30 samples. Inclusion of land use classes without a subclass may have obscured differences in spectrally heterogeneous polygons. Inclusion of land use classes not meeting SRS-stratum sample sizes may have resulted in selection of samples not truly representative of the land use class. The classified segments (selected from one of three aforementioned approaches) were processed to exclude non-agricultural land use classes and segments corresponding to known areas without vegetation (using CIR image and land use survey). The processed, CSL-classified segments were used as 1985 crop field boundaries in the modified three-tier matching.

3.4.4 Modified Three-Tier Approach to Estimate Pesticide Exposure

The Rull and Ritz (2003) three-tier approach was modified to assign PUR-derived pounds of applied pesticides for each year between 1974 and 1990 to crop fields within PLSS sections derived from the 1990 Kern County land use survey and the 1985 Landsat-classified layer (Figure 10).

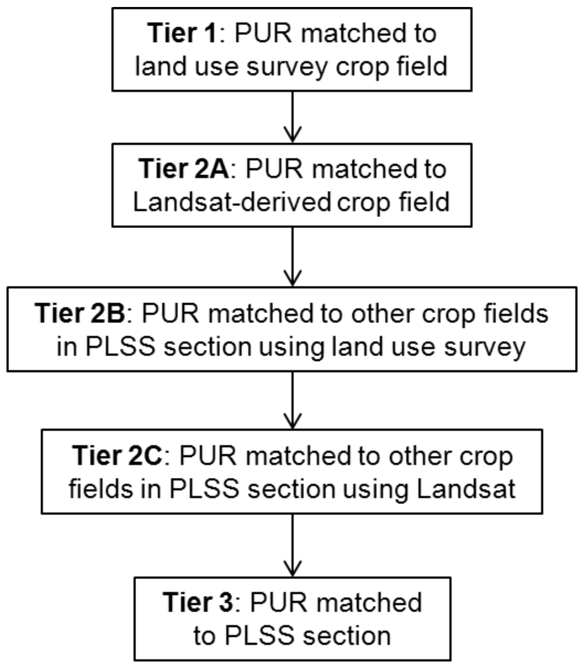


Figure 10 Modified three-tier pesticide exposure method

Land use survey and Landsat crop fields were dissolved by crop type to the PLSS section level and intersected with sections to facilitate tiered matching. Sliver polygons resulting from the intersection [area ≤0.11 ac (smallest 1990 land use survey polygon) or area ≤12.36 ac and length ≥200 m] were excluded from matching.

Tier 1 match: PUR-derived data were matched to a land use survey crop field when the crop type and PLSS section matched.

Tier 2A match: PUR data were matched to a Landsat-derived crop field when the crop type and PLSS section matched.

Tier 2B match: PUR data were matched to the other land use survey crop fields within the PLSS section.

Tier 2C match: PUR data were matched to the other Landsat-derived crop fields within the PLSS section.

Tier 3: If no land use survey and Landsat-derived crop fields were present within a PLSS section, PUR data were matched to the entire PLSS section.

The primary difference between the modified three-tier approach and the existing Rull and Ritz (2003) approach is the use of Landsat imagery to derive additional crop field information to determine likely locations of PUR applications (Tiers 2A and 2C). For each ZCTA and census tract in Kern County, organochlorine-, organophosphate-, and carbamate-specific annual pesticide application rates (lb/ac) were calculated by weighting land use survey and Landsat crop field- and section-specific application rates by the proportion of each aerial unit comprised of each crop field or section. ZCTA boundaries were clipped to the Kern County geographic extent. The weighted average of the pesticide application rates for each aerial unit was divided by 17 years to calculate an annual rate (1974 to 1990).

3.5 Rurality Metrics

The 2000 U.S. Census Bureau ZCTAs were used to approximate ZIP code boundaries; the 2000 U.S. Census Bureau census tracts were used as boundaries (U.S. Census Bureau 2013). Per Grubesic and Matisziw (2006), ZCTAs were checked for water features (HH) and large tracts of land with no mailing addresses/ZIP codes (XX). As ZCTAs may be spatially discontiguous, ZCTAs were dissolved using 5-digit ZCTA codes.

RUCA codes (2006 ZIP Code Version 2.0 and 2000 Census Tract Version) were joined to the 2000 U.S. Census Bureau ZCTA and census tract boundaries. Each ZCTA and census tract was assigned a single RUCA code. Using the recommended Categorization C, census tracts and ZCTAs with values of 1.0, 1.1, 2.0, 2.1, 3.0, 4.1, 5.1, 7.1, 8.1, or 10.1 were coded as urban; all other values were coded as rural (RHRC 2000). The 2000 version of UAs and UCs were chosen to be comparable to the RUCA codes created using 2000 U.S. Census Bureau information. The 2000 U.S. Census Bureau urban-rural classification system (U.S. Census Bureau 2000) was implemented by coding any ZCTA and census tract intersecting a UA or UC as urban. All other geographic areas were coded as rural.

3.6 Statistical Analysis

The following measures and tests quantified the accuracy of RUCA codes and the U.S. Census Bureau urban-rural system compared to the modified three-tier gold standard approach. Using 5-digit ZIP codes and census tract Federal Information Processing Standard (FIPS) codes as identifiers, the following statistical analysis was executed: For each rurality metric and pesticide chemical class, sensitivity was calculated as the number of ZCTAs, or census tracts, classified as rural divided by the number of ZCTAs, or census tracts, truly exposed to pesticides (gold standard). Specificity was calculated as the number of ZCTAs, or census tracts, classified as not rural (i.e. urban) divided by the number of ZCTAs, or census tracts, truly not exposed to pesticides (gold standard). For the gold standard, the following pesticide exposure cutoffs were evaluated: >0 lb/ac, ≥50th percentile, and ≥75th percentile of annual application rates.

Wilcoxon rank-sum tests determined if pesticide chemical class-specific annual application rates differed according to geographic aggregation (ZCTAs and census tracts) as well as if rates differed according to rurality. Separately for each areal aggregation, the kappa statistic was calculated as the proportion of the observed agreement in rurality according to each rurality metric not due to chance: (proportion of observed agreement-proportion of expected agreement due to chance) / (1-proportion of expected agreement due to chance). Separately for each rurality metric, chi-square and Fisher's exact tests determined if the proportion of ZCTAs compared to census tracts categorized as rural was different, and if the proportion of pesticide-exposed areal units differed according to the gold standard and each rurality metric. All statistical tests were two-sided (α=0.05). The data analysis was generated using the SAS System for Windows software, Version 9.3 (Copyright © 2013 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA).

CHAPTER FOUR: RESULTS

The compiled pesticide database included 157 pesticide active ingredients from three mutually exclusive pesticide chemical classes: organochlorines, organophosphates, and carbamates (Appendix A; Tables A1-A3).

4.1 PUR Extraction

There were between 476,981 and 1,305,573 PUR records each year from 1974 to 1989. Agricultural use PUR records comprised between 45.9 and 78.5% of all PUR records, representing the majority of PUR records for earlier time periods (Appendix B; Table B2). Subsequent to logic checks 1 and 2, the majority of agricultural use PUR records remained eligible for inclusion in the analysis (between 82.1% and 92.3%). As anticipated, the number of PUR records (*N*=2,657,840) available in 1990 far exceeded previous years with the adoption of full-use reporting (Appendix B; Table B3). After applying logic checks 1 through 7 to PUR records in 1990, 71.4% of agricultural use PUR records remained (Appendix B; Tables B1-B3).

Table B4 (Appendix B) shows PUR records remaining after extracting applications associated with the pesticide chemical classes of interest. Organochlorines, organophosphates, and carbamates together comprised a large proportion of agricultural use PURs - ranging between 23.2 and 48.8% between 1974 and 1990. Very few outliers were present (between 0.002 and 0.7%). Between 1974 and 1989, the majority of outliers met definition 2 (refer to Section 3.4.1). In 1990, the majority of the outliers met definition 3.

Table 5 Kern County agricultural use and chemical class PUR extractions[1]

Year	Kern County (*N*)[2]	Total (*N*)[3]
1974	11,743	11,715
1975	8,122	8,116
1976	7,228	7,219
1977	6,989	6,985
1978	8,021	8,017
1979	6,393	6,378
1980	6,423	6,419
1981	6,393	6,389
1982	6,455	6,442
1983	7,187	7,181
1984	8,207	8,195
1985	7,226	7,219
1986	9,492	9,486
1987	9,161	9,160
1988	10,898	10,894
1989	9,580	9,568
1990	19,849	19,288
Total	149,367	148,671

1 Data from CDPR (2013)
2 This number reflects PUR records after logic checks were applied and agricultural use, Kern County location, and pesticide chemical classes were extracted.
3 This number reflects PUR records additionally excluding non-agricultural and/or ambiguous PUR-derived commodities and applications reporting 0 lb.

Of the 149,367 agricultural use and chemical class-specific PUR records occurring in Kern County between 1974 and 1990, 148,671 were included in the analysis after excluding non-agricultural, ambiguous commodities (e.g. outdoor plants in containers) and 0 lb of active ingredient reported (Table 5). There were *N*=95,621 organophosphate PUR records, followed by *N*=38,436 carbamate PUR records, and *N*=14,614 organochlorine PUR records. Organophosphates were consistently the most frequently used type of chemical, followed by carbamates and organochlorines (Figures 11-12). This reflects the ban of organochlorines in the U.S. starting in 1972 - slightly

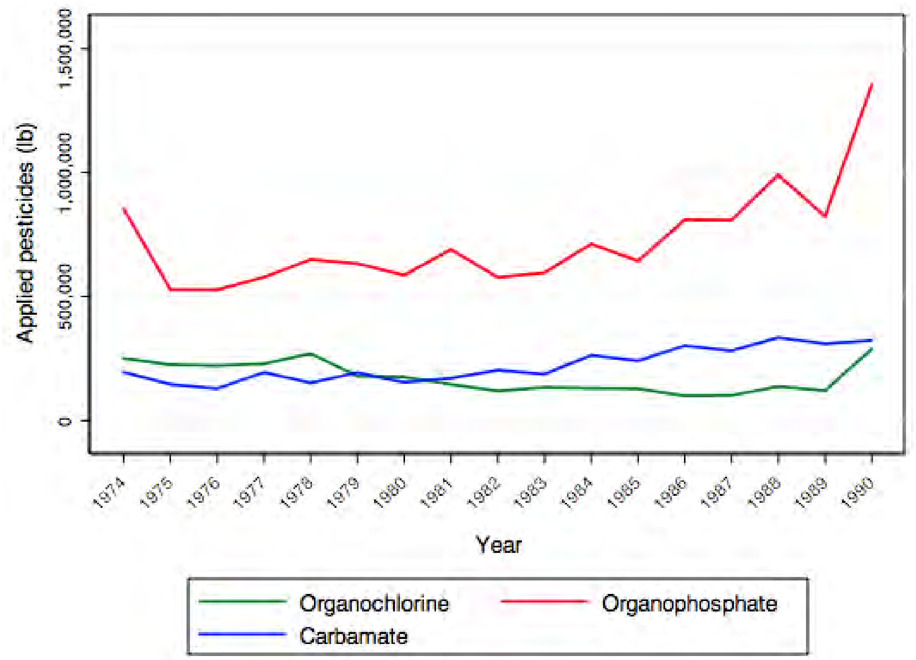


Figure 11 Pounds of agricultural pesticide usage in Kern County by chemical class (1974-1990) (Data from CDPR 2013)

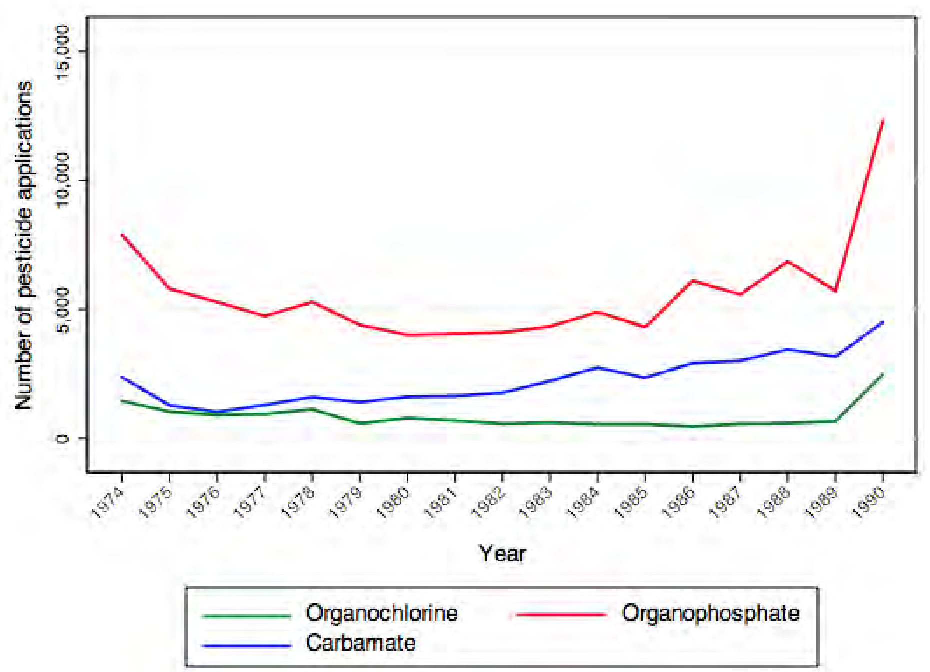


Figure 12 Agricultural PUR pesticide applications in Kern County by chemical class (1974-1990) (Data from CDPR 2013)

before PUR data became available in 1974 (CDC 2009). As demonstrated in Figures 11 and 12, the number of PUR applications mirrors the quantity of applied pesticides (lb). PUR application numbers include instances where an application was comprised of multiple active ingredients applied on a single crop type. The spike in applied pesticides and PUR applications in 1990 reflects full-use reporting beginning in 1990, where farmers were required to report all pesticide use, irrespective of restricted-use status (i.e. the pre-1990 protocol).

The most frequently pesticide-treated crops in Kern County (using PUR commodity codes) were cotton, onion, and alfalfa for organochlorines; cotton, almond, and alfalfa for organophosphates; and alfalfa, cotton, and lettuce for carbamates (Table 6). The most frequently used organochlorine, organophosphate, and carbamate pesticide active ingredient was dicofol (1.4 million lb), dimethoate (1.3 million lb), and methomyl (1.1 million lb), respectively (Table 7).

Table 6 Pesticide-treated crops by chemical class, Kern County (1974-1990)[1]

Pesticide chemical class	Crop[2]	PUR applications (*N*)
Organochlorines	Cotton	7,927
	Onion	1,711
	Alfalfa	1,511
Organophosphates	Cotton	29,302
	Almond	12,768
	Alfalfa	9,638
Carbamates	Alfalfa	7,514
	Cotton	8,083
	Lettuce	5,491

1 Data from CDPR (2013)
2 These figures use PUR commodity codes and not land use survey crop codes.

Table 7 Common pesticides by chemical class, Kern County (1974-1990)[1]

Pesticide chemical class	Pesticide active ingredient	PUR applications (*N*)	Applied chemical (lb)
Organochlorines	Dicofol	8,205	1,425,049.0
	Dacthal (DCPA)	2,023	877,093.5
	Endosulfan	1,547	169,378.7
	Methoxychlor	1,284	122,234.0
	Quintozene	550	41,285.6
Organophosphates	Dimethoate	9,658	1,263,884.6
	Tribufos	9,516	1,950,829.0
	Parathion	9,432	1,495,975.6
	Azinphos methyl	8,008	1,535,406.1
	Diazinon	6,695	803,246.7
Carbamates	Methomyl	21,041	1,061,297.3
	Aldicarb	8,001	1,046,430.0
	Carbaryl	2,406	1,072,007.0
	Carbofuran	2,314	140,738.8
	Benomyl	1,593	122,668.1

1 Data from CDPR (2013)

4.2 Crop Signature Library (CSL)

Table 8 lists the Landsat images that were processed for inclusion into the CSL. Figure 13 is a mosaic of Paths 41 and 42 and Rows 35 and 36 created from radiometrically and atmospherically corrected October 1990 band 3 (red) images subsequent to applying a median spatial filter, reclassifying negative reflectance values to 0, and cropping to a geographic extent enclosing Kern County. Figure 14 allows for a larger scale examination of the crop fields captured by Landsat imagery. Figures 15 and 16 show the aforementioned processing, but using band 4 (near infrared) Landsat images from October 1990. Original mosaics for the bands 3 and 4 images are shown in Appendix C (Figures C1-C2).

Table 8 Landsat images from 1990 used for crop signature library[1,2]

Month		Path 41		Path 42	
		Row 35	**Row 36**	**Row 35**	**Row 36**
January	Cloud cover	10%	0%	0%	0%
	Acquisition	1/22/1990	1/22/1990	1/29/1990	1/29/1990
February	Cloud cover	0%	0%	0%	0%
	Acquisition	2/15/1990	2/15/1990	2/14/1990	2/14/1990
March	Cloud cover	10%	Excluded	0%	10%
	Acquisition	3/27/1990		3/18/1990	3/18/1990
April	Cloud cover	40%	20%	0%	40%
	Acquisition	4/28/1990	4/28/1990	4/3/1990	4/3/1990
May	Cloud cover	10%	10%	0%	10%
	Acquisition	5/30/1990	5/30/1990	5/5/1990	5/5/1990
June	Cloud cover	Excluded	Excluded	0%	10%
	Acquisition			6/6/1990	6/6/1990
July	Cloud cover	None available	None available	0%	50%
	Acquisition			7/8/1990	7/8/1990
August	Cloud cover	10%	10%	0%	0%
	Acquisition	8/18/1990	8/18/1990	8/25/1990	8/25/1990
September	Cloud cover	0%	0%	0%	20%
	Acquisition	9/3/1990	9/3/1990	9/10/1990	9/10/1990
October	Cloud cover	0%	0%	0%	0%
	Acquisition	10/5/1990	10/5/1990	10/12/1990	10/28/1990

1 Data from USGS (2013b)

2 Landsat images were not available for November and December 1990. Images were excluded due to excessive cloud cover overlapping the Kern County geographic extent.

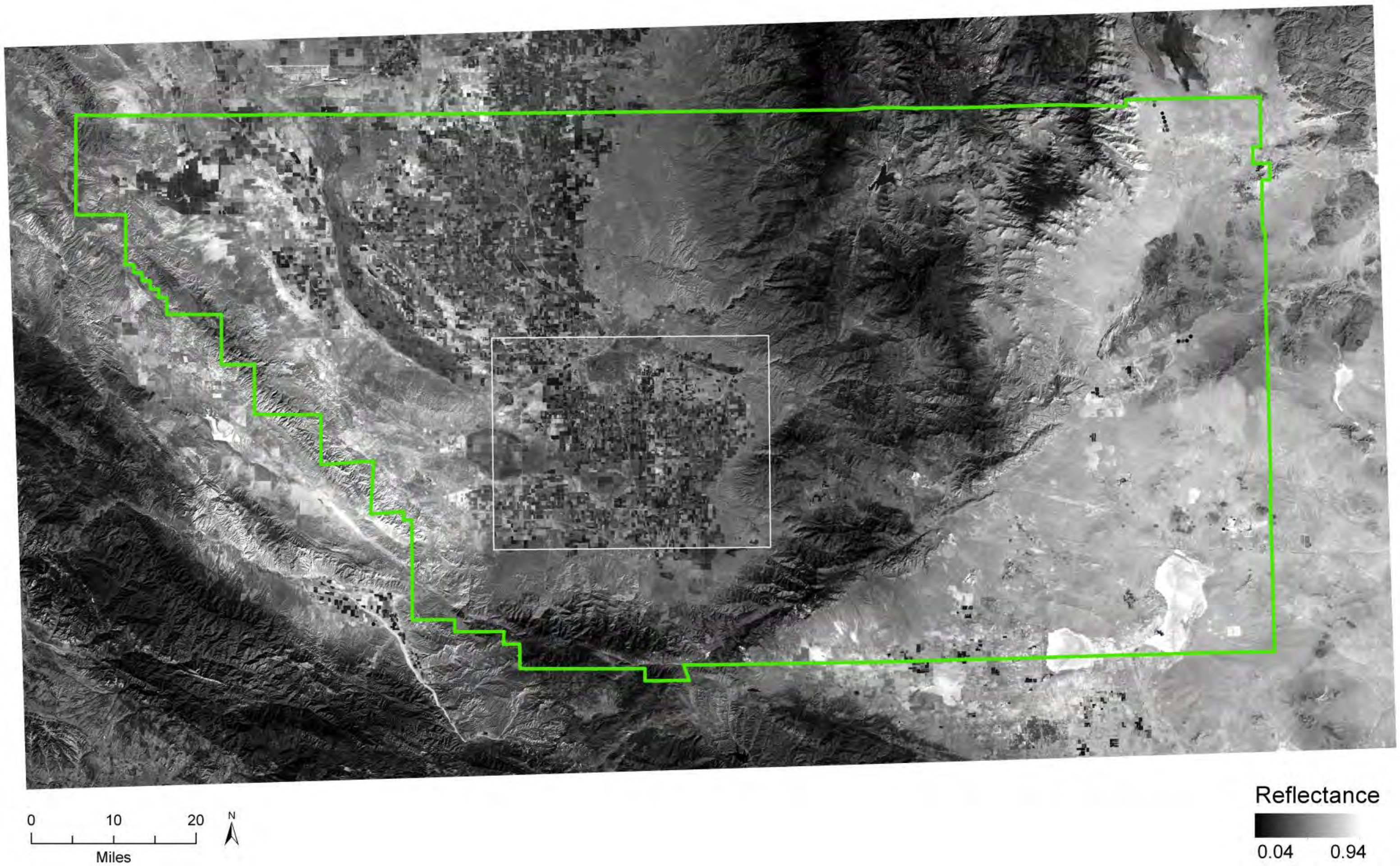


Figure 13 Landsat mosaic (band 3), Paths 41-42 and Rows 35-36, from October 1990 cropped to Kern County (Data from U.S. Census Bureau 2013; and USGS 2013b)

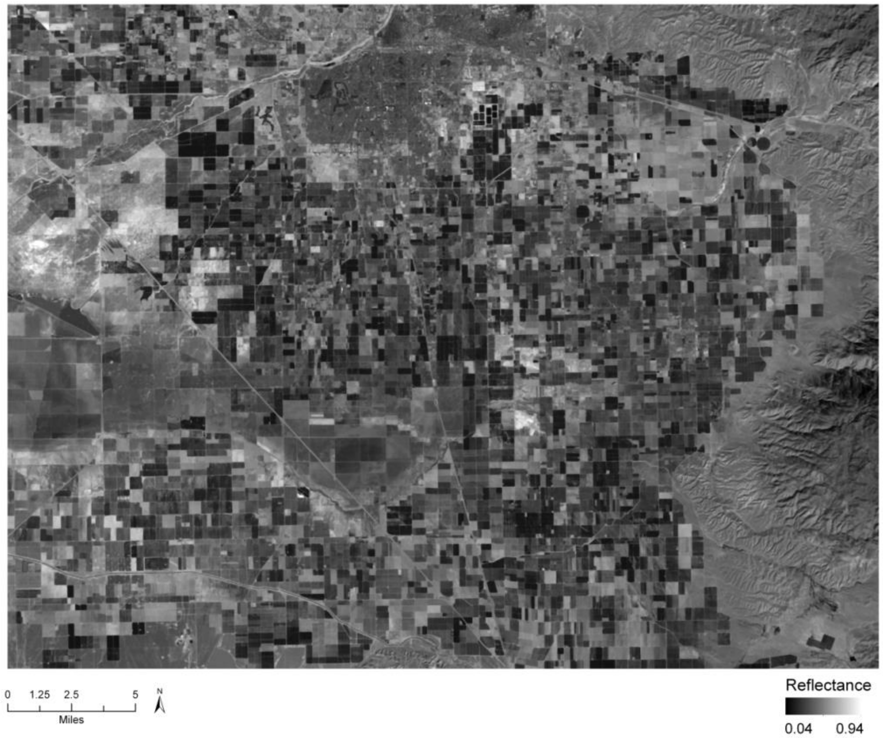


Figure 14 Inset of Landsat mosaic (band 3) from October 1990, showing crop fields in Kern County (Data from USGS 2013b)

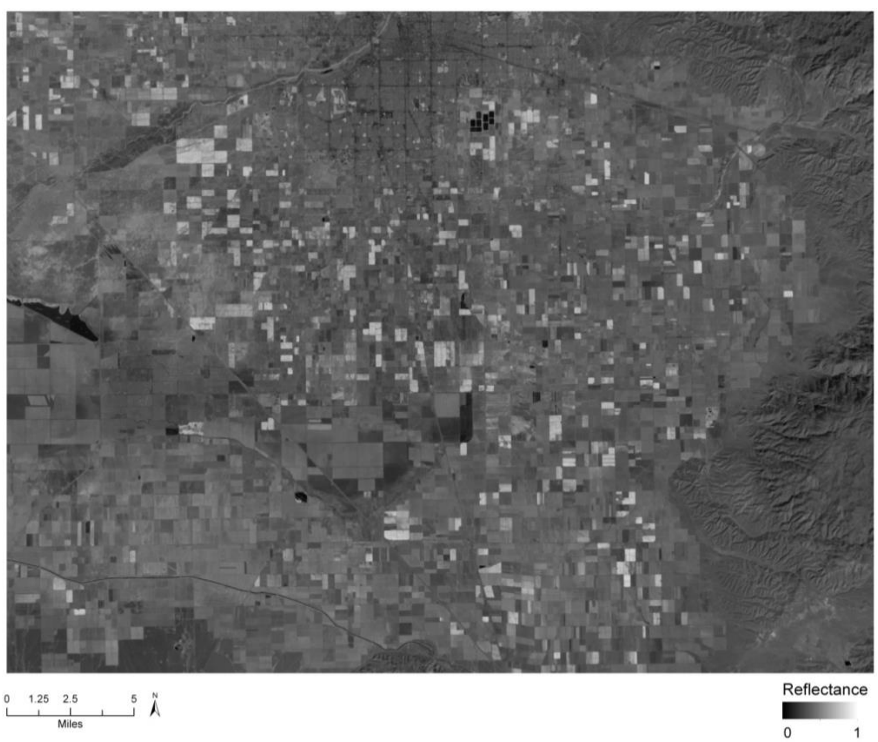


Figure 15 Inset of Landsat mosaic (band 4) from October 1990, showing crop fields in Kern County (Data from USGS 2013b)

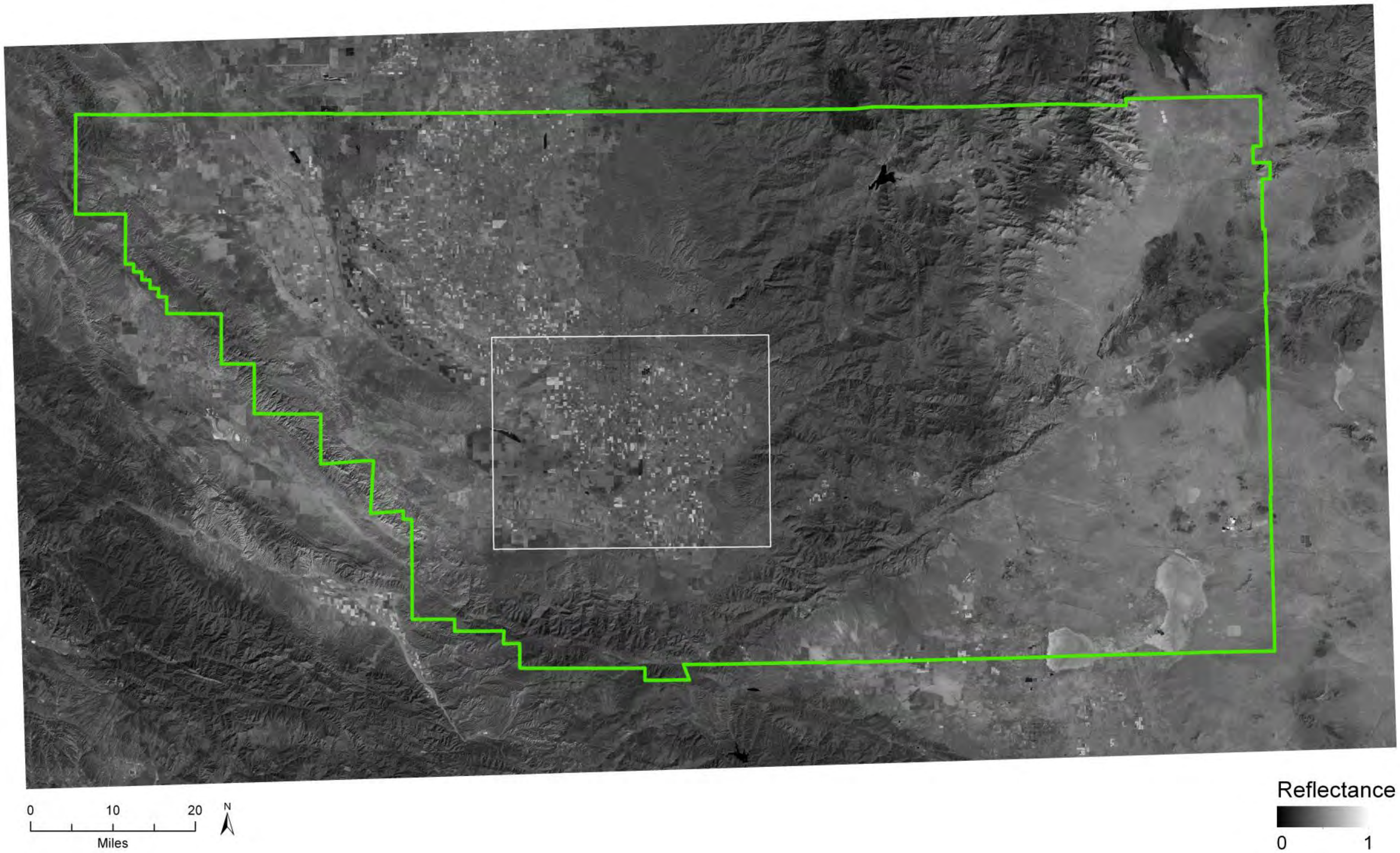


Figure 16 Landsat mosaic (band 4), Paths 41-42 and Rows 35-36, from October 1990 cropped to Kern County (Data from U.S. Census Bureau 2013; and USGS 2013b)

NDVI images were created using the red and near infrared bands for each month between January and October 1990. For example, in October 1990, NDVI values ranged between -0.52 and 1; negative values are indicative of non-vegetation (e.g. barren rock) and positive values closer to 1 are indicative of dense vegetation (Figures 17 and 18).

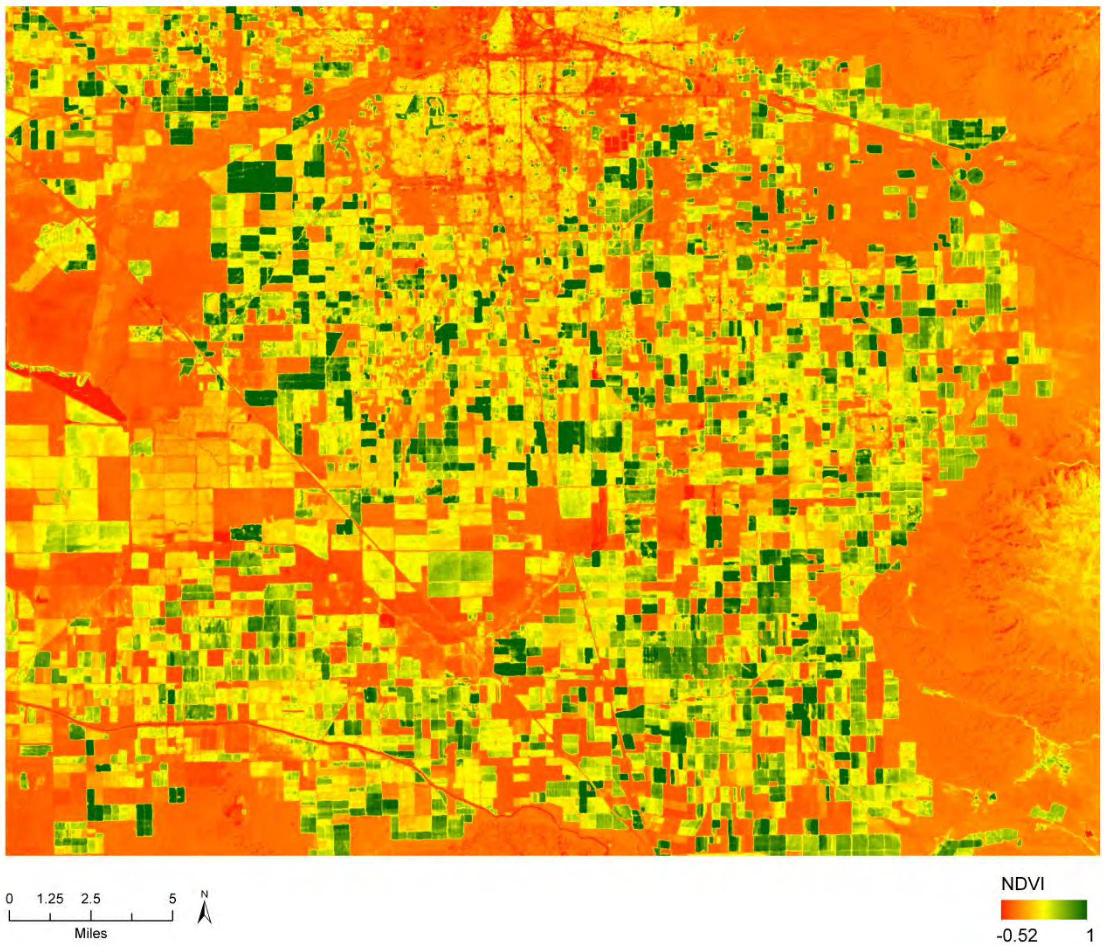


Figure 17 Inset of NDVI image created from red and near infrared Landsat bands, October 1990 (Data from USGS 2013b)

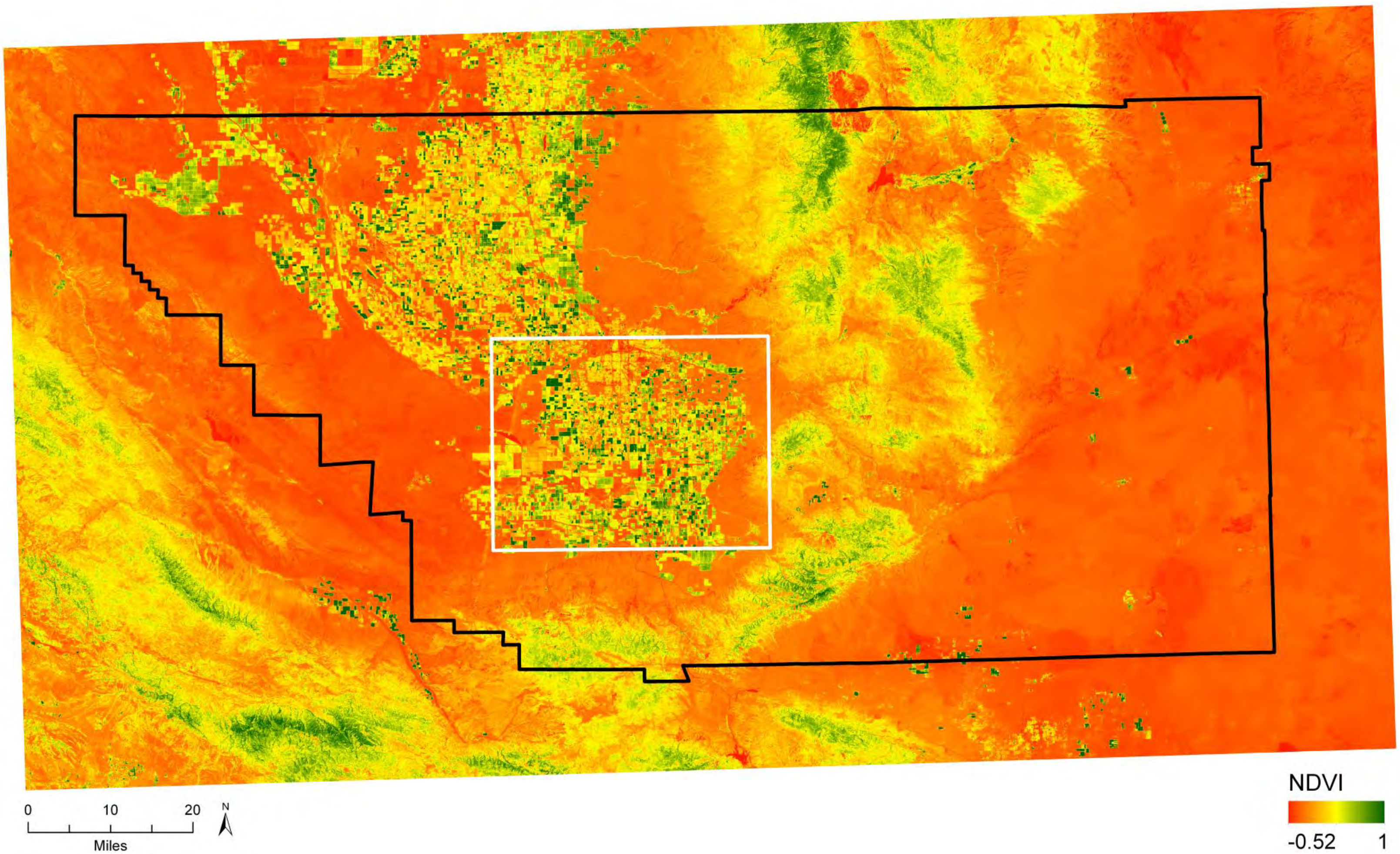


Figure 18 NDVI image cropped to Kern County, October 1990
(Data from U.S. Census Bureau 2013; USGS 2013b)

4.2.1 Stratified Random Sampling (SRS)

Stratified random sampling (SRS) was applied to the 1990 Kern County land use survey in geographic areas meeting three eligibility criteria. SRS-eligible land use survey polygons must have been: (1) single-use; (2) at least 4 ha in area; and (3) within the cloud-free zone (Table 9). A total of 16,635 of the 16,769 land use survey polygons included in the 1990 Kern County dataset satisfied the single-use criterion, followed by 12,197 satisfying the single-use and area criteria, and 11,832 satisfying all three criteria. The majority of crop types excluded from 1990 Kern County land use survey due to double-cropping or intercropping/mixed (i.e. not single-use) were grain and hay crops (N=55; 41% non-single-use), almonds (N=33; 24.6%), and potatoes (N=29; 21.6%) (Table 10). Figure 19 shows the cloud-free zone representing the geographic area where cloud- and shadow-free NDVI images were available for all months in 1990. As July 1990 was missing Path 41, this portion of the study area was excluded.

Out of the 11,832 land use survey polygons eligible for SRS, 1,423 were randomly selected within 81 land use strata (excluded Z: outside of study area) (Appendix D; Table D1). At most 30 samples were randomly selected within each stratum (Figure 20). However, 49 land use classes had samples sizes less than 30 due to a low prevalence of such classes in Kern County subsequent to applying the aforementioned eligibility criteria. SRS samples for these land use classes may not be representative of the strata.

Table 9 Eligibility criteria for SRS[1]

Eligibility criteria	**Sample (*n*)[a]**
Single-use	16,635
≥4 ha	12,197
In cloud-free zone	11,832

[1] Data from CDWR (2013)

Table 10 Land use classes excluded from SRS due to multiuse[1]

Land use class	***N* (%)**
Grain and hay crop	55 (41.0%)
Almond	33 (24.6%)
Potato	29 (21.6%)
Onion and garlic	7 (5.2%)
Corn	3 (2.2%)
Carrot	2 (1.5%)
Melon, squash, cucumber	2 (1.5%)
Cotton	1 (0.8%)
Pistachio	1 (0.8%)
Tomato	1 (0.8%)

[1] Data from CDWR (2013)

NDVI values for each pixel of each NDVI image between January and October 1990 intersecting any of the SRS-sampled land use survey polygons were extracted. A total of 645,127 NDVI values were extracted for each month - from a total of 6,451,270 NDVI values contributing to the 1990 CSL (Appendix D; Table D2). This was subsequent to removing one pixel with an NDVI value associated with a point that straddled the line between two land use survey polygons. For any given month between January and October 1990, the native vegetation land use class contributed the most NDVI values to the CSL (*N*=149,648) across all of its 30 SRS-sampled polygons. This reflects the typically large size of its polygons (median 71.75 ac vs. 43.44 ac for all other

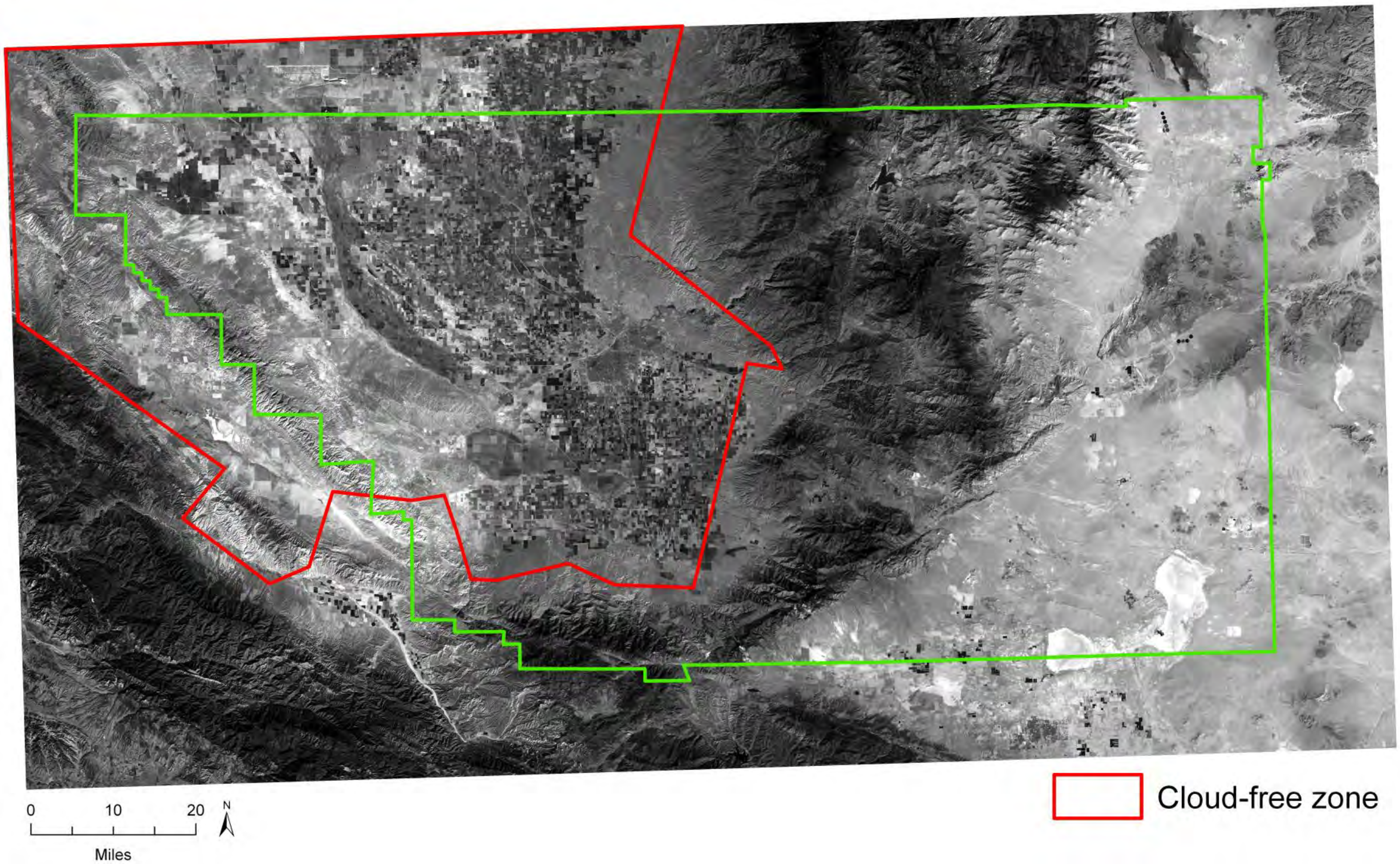


Figure 19 Cloud-free zone of 1990 Landsat images available for CSL
(Data from U.S. Census Bureau 2013; and USGS 2013)

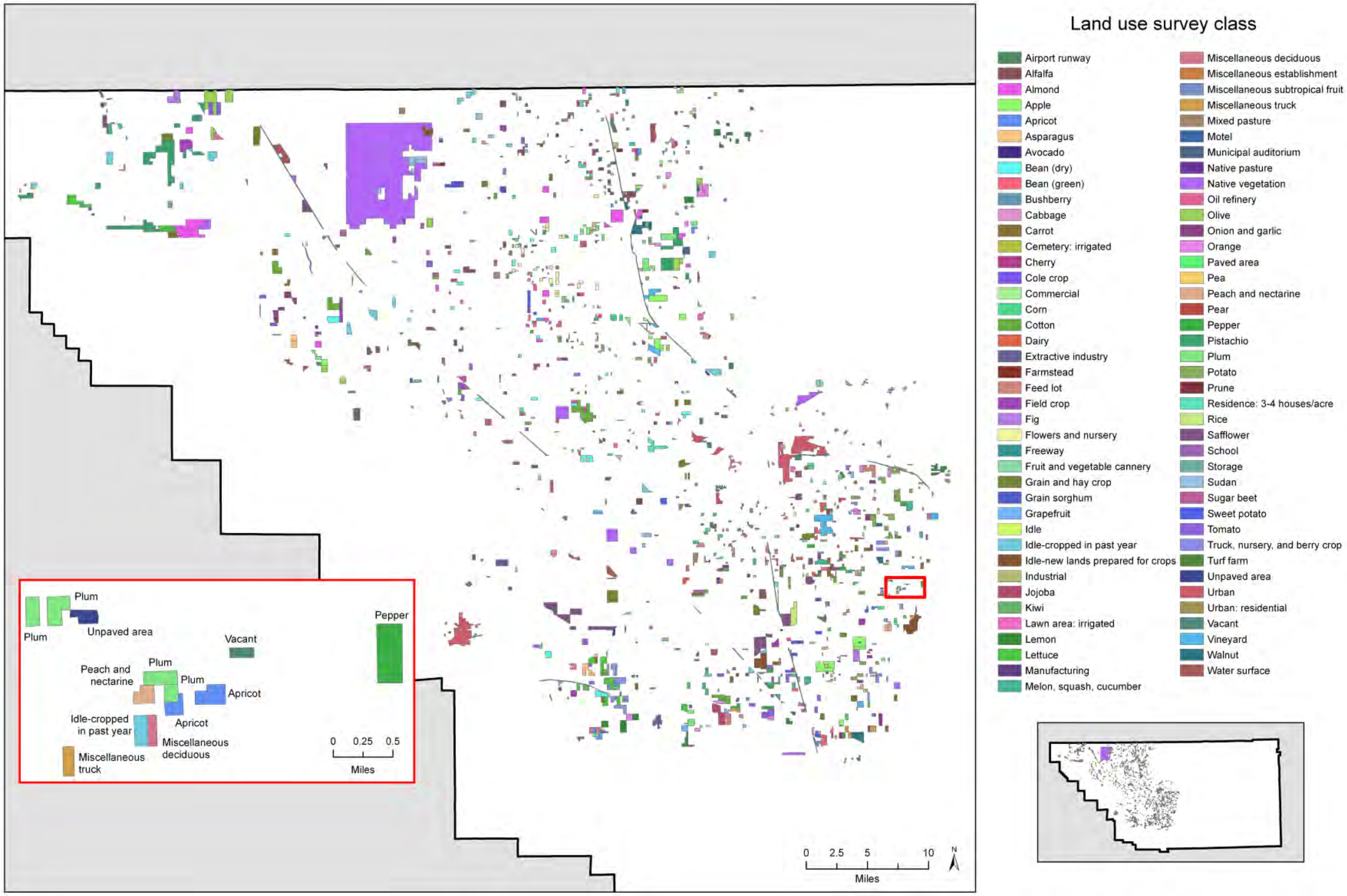


Figure 20 Land use survey polygons sampled via SRS, Kern County, 1990
(Data from CDWR 2013; and U.S. Census Bureau 2013)

land use classes). Commercial (motel) contributed the fewest NDVI values (*n*=1 SRS; *N*=52 NDVI). Out of the agricultural land use classes of interest to the analysis, pistachio contributed the most NDVI values (*n*=30 SRS; *N*=26,878 NDVI) and avocado contributed the fewest NDVI values (*n*=1 SRS; *N*=61 NDVI).

Figure 21 shows select SRS-sampled land use survey polygons and their extracted NDVI pixel values for October 1990. Median NDVI values for the selected peach and

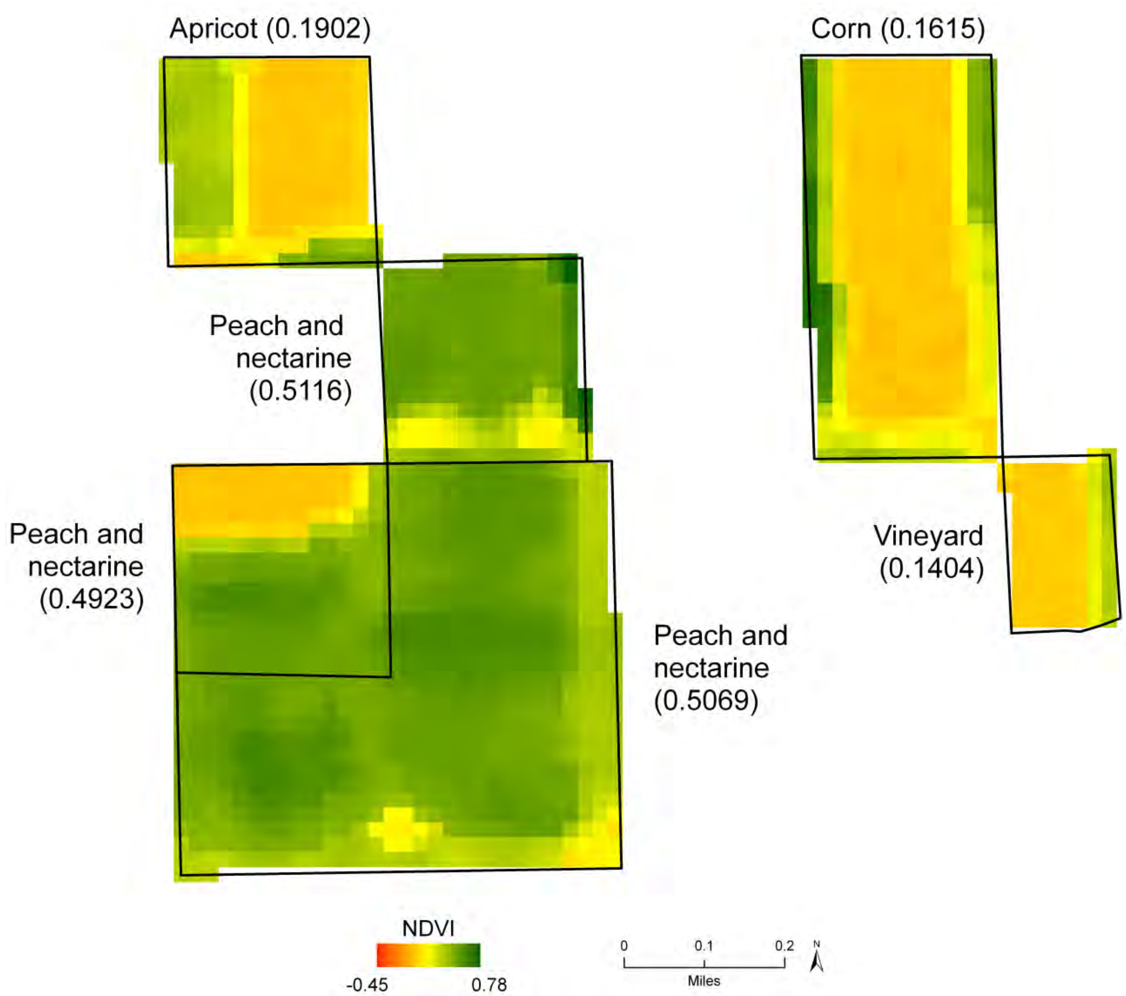


Figure 21 Median NDVI values for select SRS-sampled land use survey polygons, October 1990. The polygons are located near PLSS section 15M32S29E14. (Data from CDWR 2013)

nectarine land use survey polygons are similar (ranging from 0.4923 to 0.5116). Variability between different land use survey polygons is manifest in the low median NDVI value of the vineyard land use survey polygon (0.1404) relative to the highest median NDVI value of the peach and nectarine land use survey polygons (0.5116).

As an objective improvement over the Maxwell (2011) approach, all NDVI values of each SRS-sampled polygon were used to compute a median NDVI value for that specific polygon - harnessing all available spectral data from the NDVI images. This is in contrast to the Maxwell (2011) method, which selects one pixel per polygon. The final CSL contained median NDVI values for all SRS-sampled polygons from each land use class (using the median NDVI value for each polygon) from January and October 1990. All negative NDVI values were recoded to 0 (no vegetation) (Beck et al. 2006). Refer to Appendix D (Figures D1-D56) for figures showing NDVI values of all agricultural land use classes included in the CSL.

4.3 Classification of 1985 Landsat Imagery

4.3.1 Segmentation

Table 11 lists the Landsat images used to create a classified 1985 Kern County crop field layer using the CSL. February was excluded due to the absence of the majority of agricultural crop fields (Path 42) and November and December were not considered as the CSL only extended into October. A Landsat Multispectral Scanner (MSS) image was available for July 1985, but not used due to a different spatial resolution compared to Thematic Mapper (TM) images. Subsequent to radiometric and atmospheric processing, mosaicking, reclassification, spatial filtering, cropping to Kern County, and NDVI

Table 11 Landsat images from 1985 used for classification[1,2]

Month		Path 41		Path 42	
		Row 35	Row 36	Row 35	Row 36
January	Cloud cover			20%	0%
	Acquisition	Excluded	Excluded	1/31/1985	1/31/1985
February	Cloud cover	10%	10%	None available	None available
	Acquisition	2/25/1985	2/25/1985		
March	Cloud cover			0%	0%
	Acquisition	Excluded	Excluded	3/20/1985	3/20/1985
April	Cloud cover	10%	10%	0%	50%
	Acquisition	4/14/1985	4/14/1985	4/5/1985	4/5/1985
May	Cloud cover	10%	10%	0%	20%
	Acquisition	5/16/1985	5/16/1985	5/23/1985	5/23/1985
June	Cloud cover	10%	1%	0%	18%
	Acquisition	6/1/1985	6/17/1985	6/8/1985	6/8/1985
July	Cloud cover	0%	0%	1%	None available
	Acquisition	7/3/1985	7/3/1985	7/26/1985	
August	Cloud cover	None available	10%	0%	50%
	Acquisition		8/20/1985	8/11/1985	8/11/1985
September	Cloud cover	0%	0%	1%	0%
	Acquisition	9/21/1985	9/21/1985	9/12/1985	9/12/1985
October	Cloud cover			0%	0%
	Acquisition	Excluded	Excluded	10/14/1985	10/14/1985

1 Data from USGS (2013b)
2 Images were excluded due to excessive cloud cover overlapping the Kern County geographic extent

calculations, all NDVI images were clipped to the geographic extent containing NDVI values for all months in January and March through October and containing no clouds. Figure 22 shows the geographic extent of the segmentation-eligible zone (no clouds and available 1985 data) compared to the cloud-free zone used to determine eligible SRS land use survey polygons. Note that some reflectance values exceed 1, which is physically acceptable and may arise from bright surfaces (e.g. clouds) (YCEO 2013).

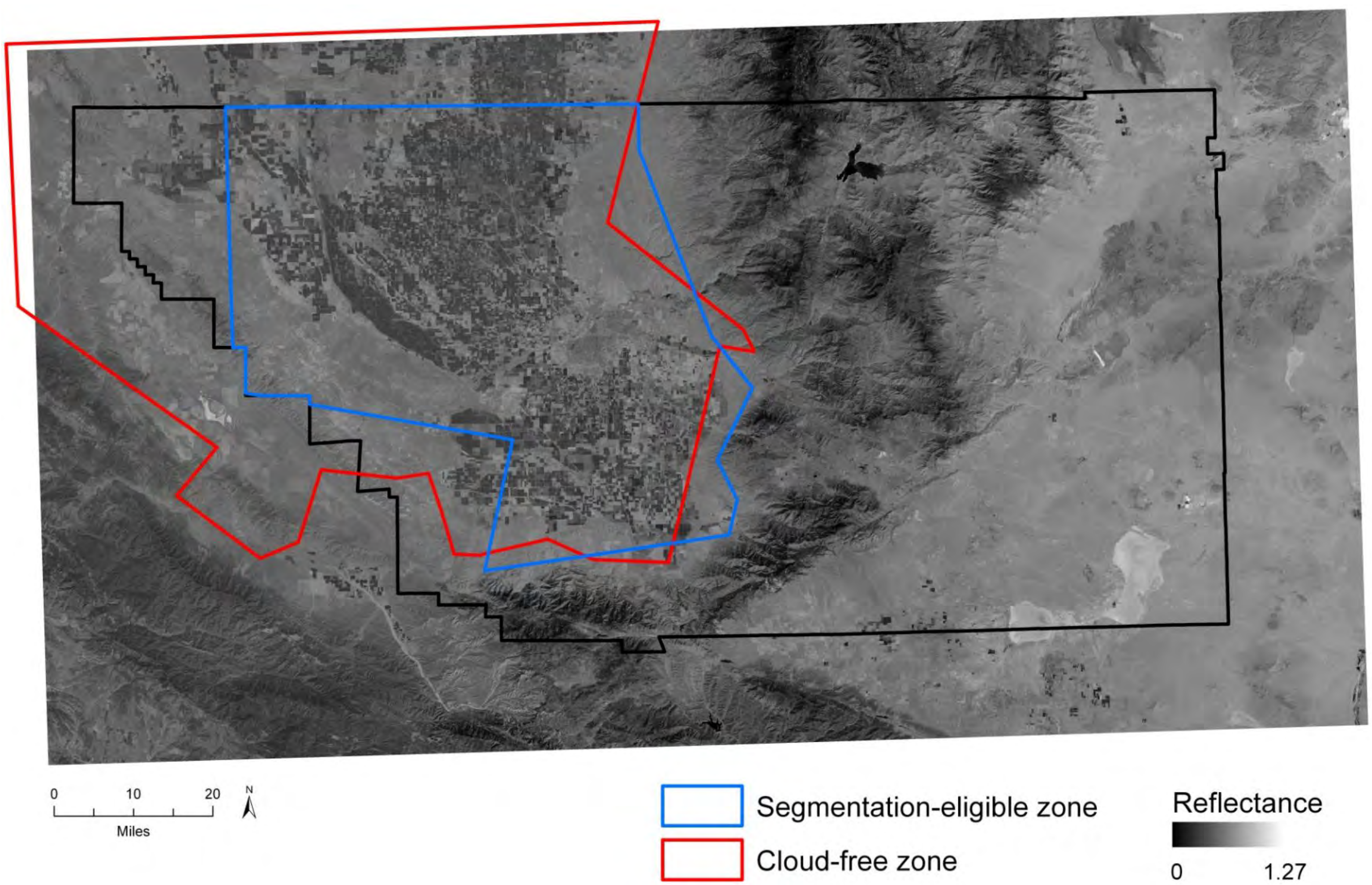


Figure 22 Segmentation-eligible zone vs. cloud-free CSL zone, overlaying Landsat mosaic (band 3) from September 1985 (Data from USGS 2013b; and U.S. Census Bureau 2013)

4.3.2 Principal Component Analysis (PCA)

A principal component analysis (PCA) was performed using the nine NDVI images (January, March to October 1985) (Table 12). Over 82% of the overall variance is explained by the first three components. These three principal components were used to create a crop field boundary layer via segmentation (Appendix E; Figure E1). The segmented polygon feature class consisted of 19,752 segments, each representing a spectrally homogeneous grouping of pixels derived from the three input principal components (Figures 23 and 24).

Table 12 Principal component analysis of Landsat 1985 NDVI images

Principal component	Variance (%)	Principal component	Variance (%)
1	56.06	6	3.15
2	16.84	7	2.82
3	9.42	8	2.40
4	4.63	9	1.13
5	3.54		

4.3.3 Classification Using Sum of Squared Difference (SSD)

NDVI values for each segment were extracted from monthly NDVI images in 1985. There were a total of 7,825,045 NDVI values across all segments for each month. There was an average of 396 NDVI values intersecting each segment (median 258, minimum 19, maximum 7,742), and thus contributing to the classification of each segment. Median NDVI values for each segment were compared to median NDVI values from each land use class in the CSL using a sum of squared differences (SSD) measure. Each segment was classified by assigning it to one land use class based on the smallest SSD. Three

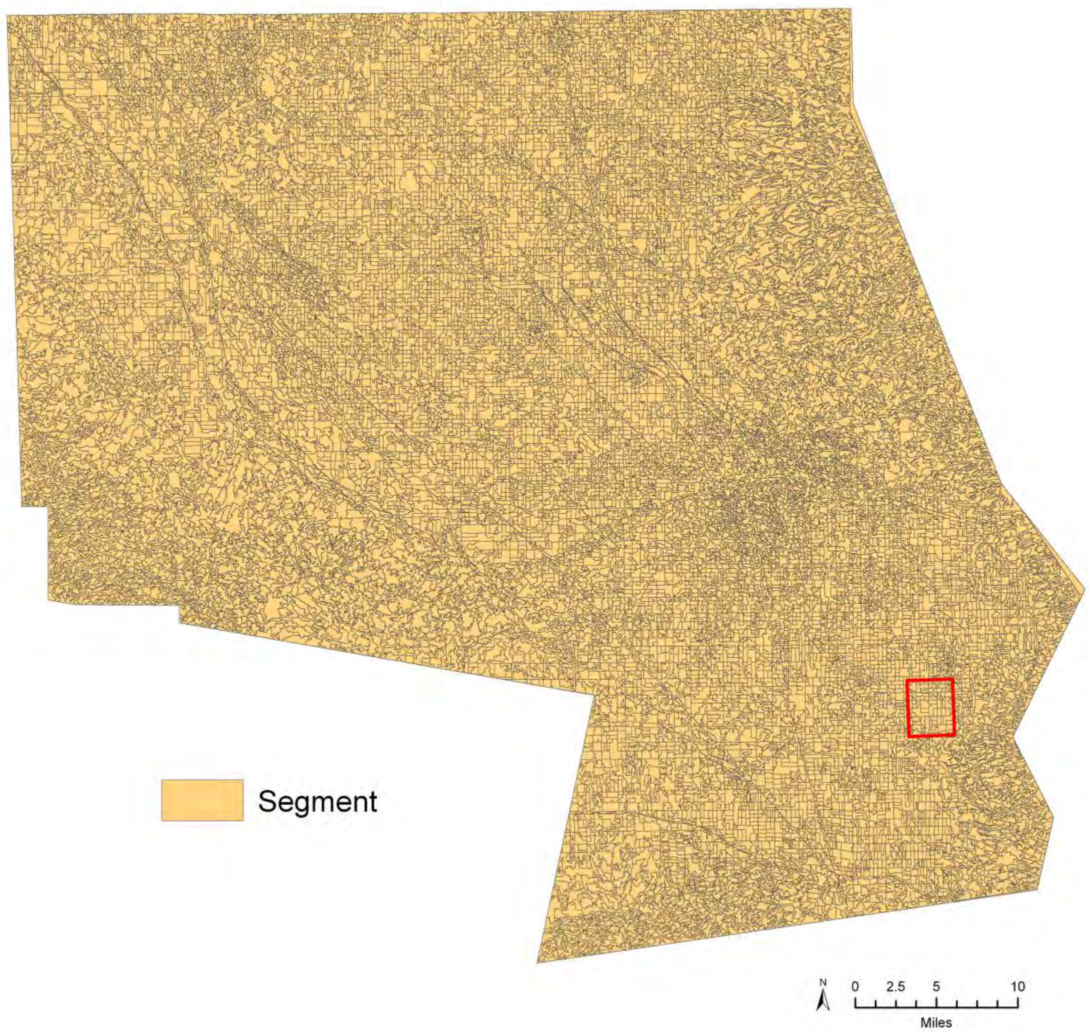


Figure 23 Segments of spectrally homogeneous pixels, basis of crop field boundaries for classifying 1985 Landsat NDVI images

different CSL classification approaches were executed, each more conservative than the previous (Table 13). Minimum SSDs did not differ between classifications 1 and 2 (median of 0.06) (Table 14). The minimum SSD of classification 3 (median of 0.08) was slightly different from classifications 1 and 2 - indicating a larger difference in NDVI values for land use matches using classification 3.

Figure 24 Segments overlaying color-infrared Landsat image
from August 1985 (Data from USGS 2013b)

Table 13 CSL classification approaches for segmented crop layer

Classification	Description	Land use classes (*N*)
1: Standard	Excluded Z land use class (outside of study area)	81
2: Subclass-required	Excluded land use classes without specified subclass	75
3: Strict	Excluded land use classes without specified subclass and with SRS samples <30	28

Table 14 Classification: minimum sum of squared differences (SSD)

	N	Mean ± *SD*	Median (*IQR*)	Min.	Max.
Classification 1	19,752	0.09 ± 0.09	0.06 (0.10)	0.001	0.84
Classification 2[1]	19,752	0.09 ± 0.09	0.06 (0.11)	0.001	0.84
Classification 3	19,752	0.12 ± 0.12	0.08 (0.13)	0.002	0.87

[1] Classification 2 (subclass-required) was selected to classify 1985 Landsat images.

However, the crop types with the largest SSDs did vary according to classification method (using 99th percentile cutoff). For classification 1 among segments with SSDs ≥0.39 NDVI, apricots comprised the majority (*N*=21; 10.6%) (data not shown). For classification 2 among segments with SSDs ≥0.40 NDVI, cole crops comprised the majority (*N*=30; 15.2%). For classification 3 among segments with SSDs ≥0.54 NDVI, oranges comprised the majority (*N*=54; 27.4%). Results may differ if using land use class-specific SSD percentiles.

In terms of the 20 most frequently classified land use classes (Appendix E; Figures E2-E4), classifications 1 and 2 were similar, as a large number of segments were classified as jojoba (classification 1: *N*=2,308; classification 2: *N*=2,618) and cotton (classification 1: *N*=1,879; classification 2: *N*=1,879). Classification 3 was dramatically different from classifications 1 and 2, as most segments were classified as feed lots (*N*=3,335), a semi-agricultural land use class.

After comparing and contrasting the three classified crop field layers, classification 2 (subclass-required) was chosen for its comparability to classification 1 in land use classification frequencies, to address the potential heterogeneity in including land use classes without a specified subclass, and the decision to not break the SRS randomization used in creating the CSL. Figure 25 is a map showing the preprocessed classified crop field layer (using classification 2), which is dominated by jojoba crop fields along the periphery. Cotton fields displayed in blue are interspersed throughout the layer.

4.3.4 Processing CSL-Classified Crop Fields

The original classified layer (Figure 25) was iteratively processed to exclude non-agricultural segments using the 1990 Kern County land use survey and a color-infrared (CIR) band combination of Landsat images in August 1985. Using the CSL-classified land use classes assigned to the segments, non-agricultural classes were deleted (e.g. urban). There was a potential misclassification of segments assigned to the jojoba land use class. A large number of segments were classified as jojoba (*N*=1,572 after aforementioned processing). However, there were few jojoba land use survey polygons according to the 1990 Kern County land use survey (*N*=9). It is conceivable that jojoba constituted some of the native vegetation land use classes, commonly present in the same peripheral geographic regions of the land use survey. The discrepancy between the relatively few jojoba crop fields in Kern County in 1990, a short time after the 1985

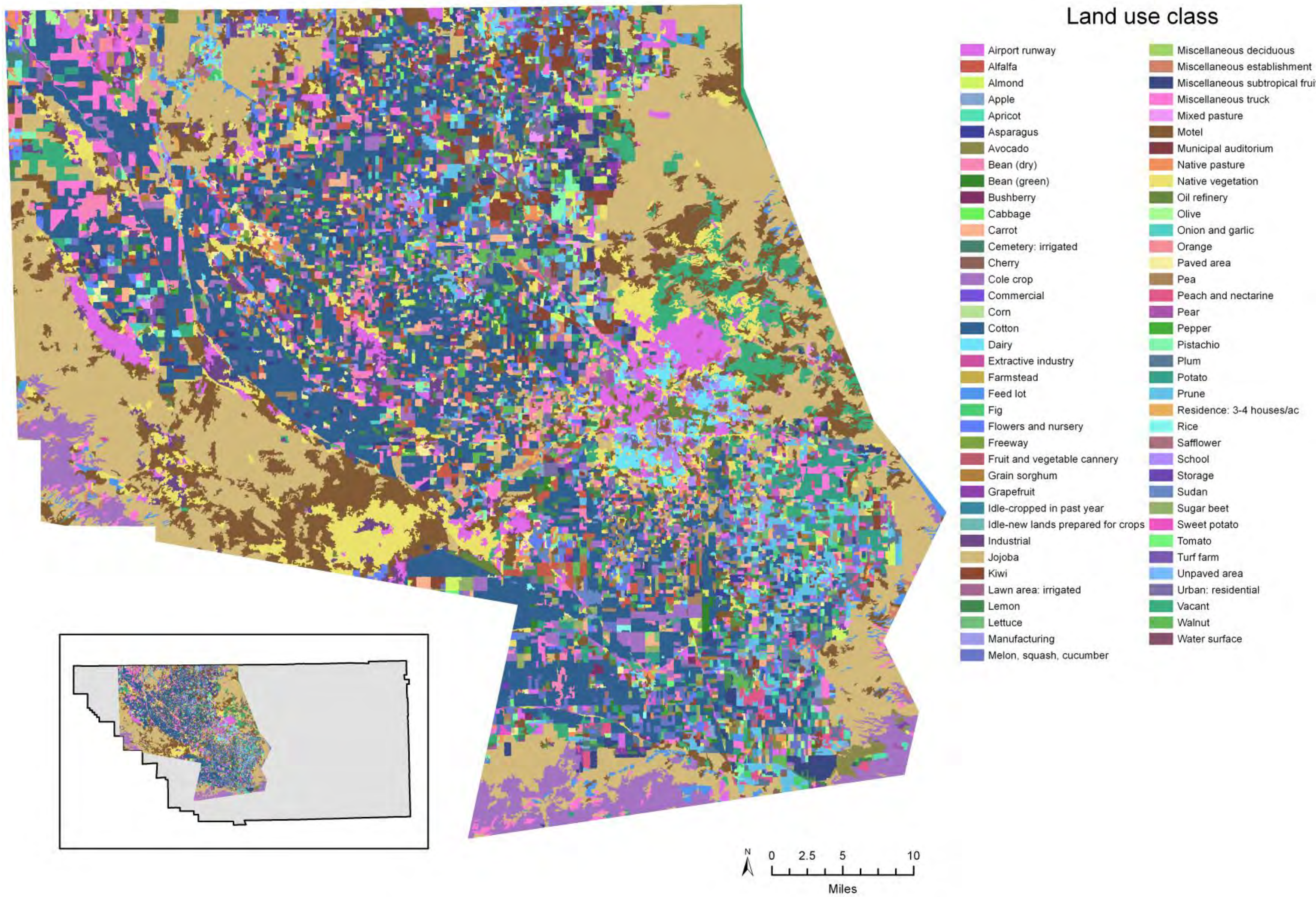


Figure 25 Classification 2-derived segments prior to processing

Landsat images were acquired (used for CSL classification), lends support to likely misclassification. These jojoba-classified segments potentially share a similar spectral profile between the months of January and March to October with some other land use class either included in or absent from the CSL (Appendix D; Figure D26). All jojoba segments were excluded from the analysis.

Figure 26 shows the final CSL-classified 1985 Kern County crop field boundaries including 10,008 crop fields. The percentages of agricultural use polygons in the 1990 land use survey (Appendix E; Table E1) slightly differ from the CSL-classified layer (Appendix E; Table E2). Note that the land use survey geographic extent is larger than that of the CSL-classified layer. There was a total of 49 land use classes in the 1985 classified layer - predominantly cotton (N=1,878; 18.76%). This number overestimates the actual number of cotton fields as single cotton fields may have been represented as multiple adjacent segments by virtue of the segmentation process.

4.4 Modified Three-Tier Approach

Processed PUR pesticide records (lb of applied AI) for every year between 1974 and 1990 were matched to the 1990 Kern County land use survey, the newly created CSL-classified crop field layer derived from 1985 Landsat data, or PLSS section data (Figures 27-29). Across all three chemical classes, the majority of PUR records matched to Tier 1: 84.5% of organochlorine records, 85.5% of organophosphate records, and 83% of carbamate records. The contribution of the CSL-classified Landsat layer was modest, ranging from 2.1 to 2.4% for Tier 2A and from 0.1 to 0.2% for Tier 2C. Had the land use survey not been used as the first tier, more PUR records would likely match to the

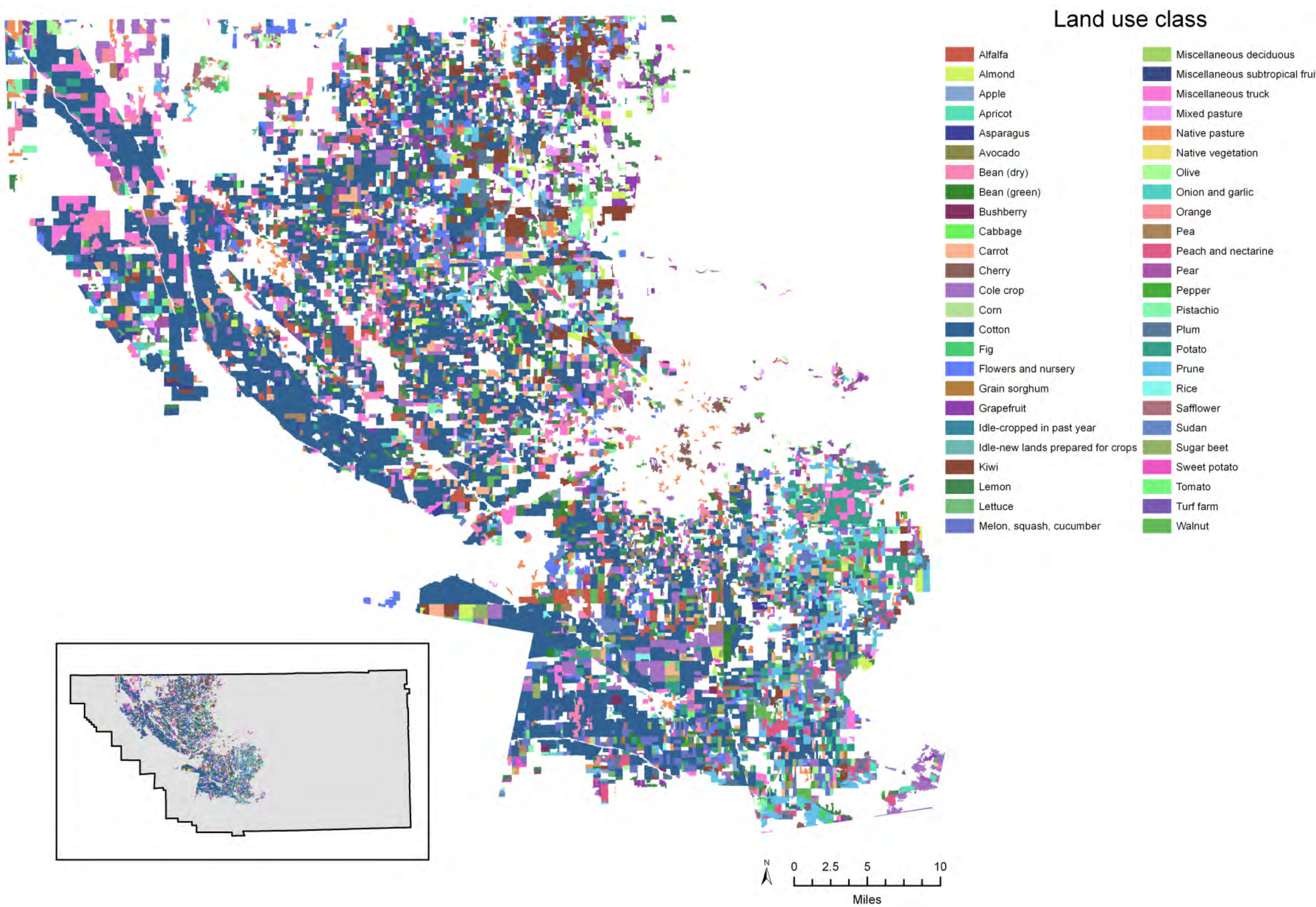


Figure 26 Finalized classification 2-derived segments subsequent to processing

Landsat-derived crop fields - particularly PUR records of applied pesticides close in time to 1985. Percentage tier matches may also differ when examining individual years. Note that the number of PUR records reflects the application of different individual pesticide active ingredients. Some records may also be a part of the same pesticide application, comprised of multiple pesticides ingredients applied on crop.

The PUR records for each chemical class between 1974 and 1990 are associated with 2,952,761.16 lb (*N*=14,614 PUR records) for organochlorines, 12,367,594.88 lb (*N*=95,621) for organophosphates, and 3,777,562.21 lb (*N*=38,436) for carbamates. Note that these totals reflect rounding error inherent in distributing PUR pounds to all crop fields in a PLSS section via Tiers 2B and 2C.

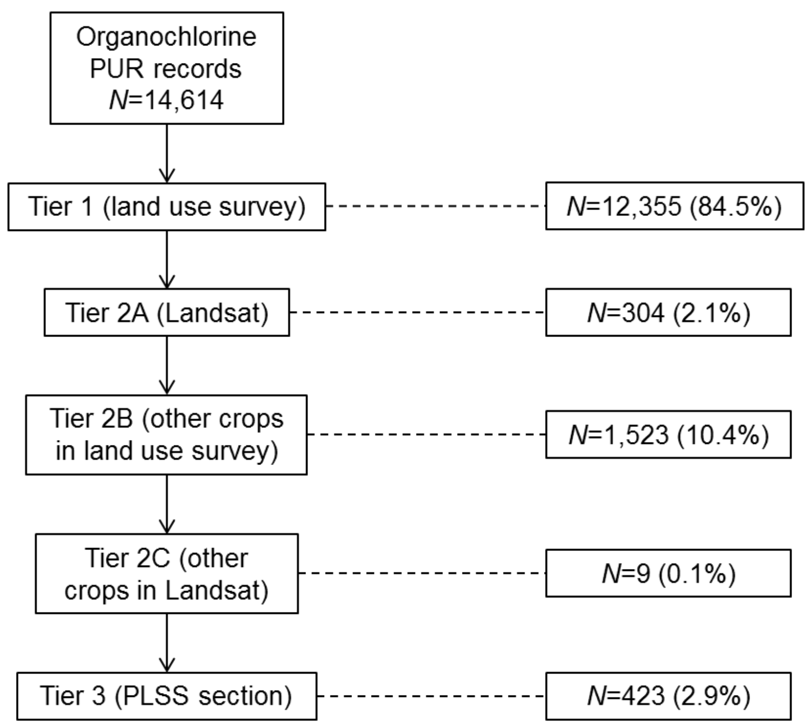


Figure 27 Organochlorine PUR tier matches, Kern County (1974-1990)

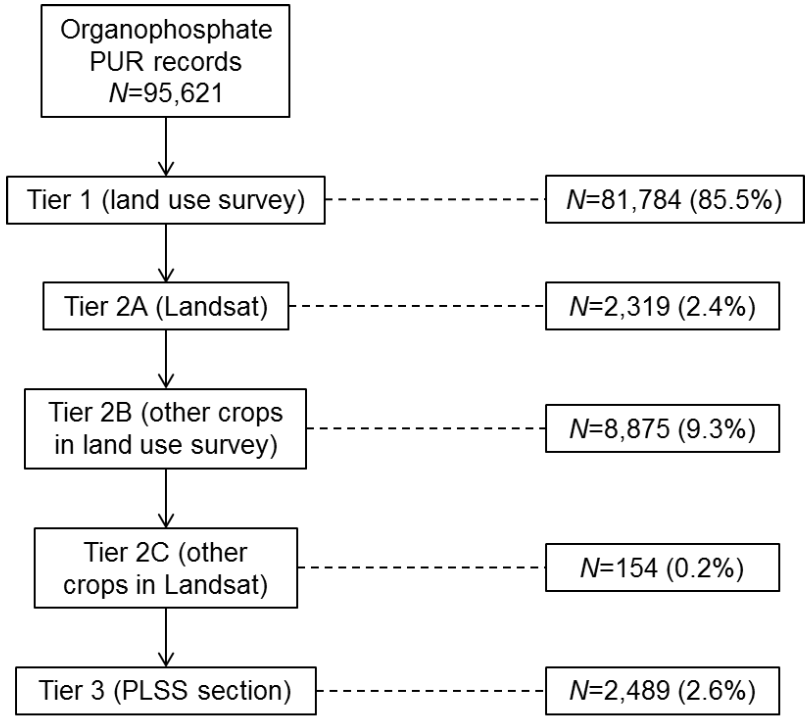


Figure 28 Organophosphate PUR tier matches, Kern County (1974-1990)

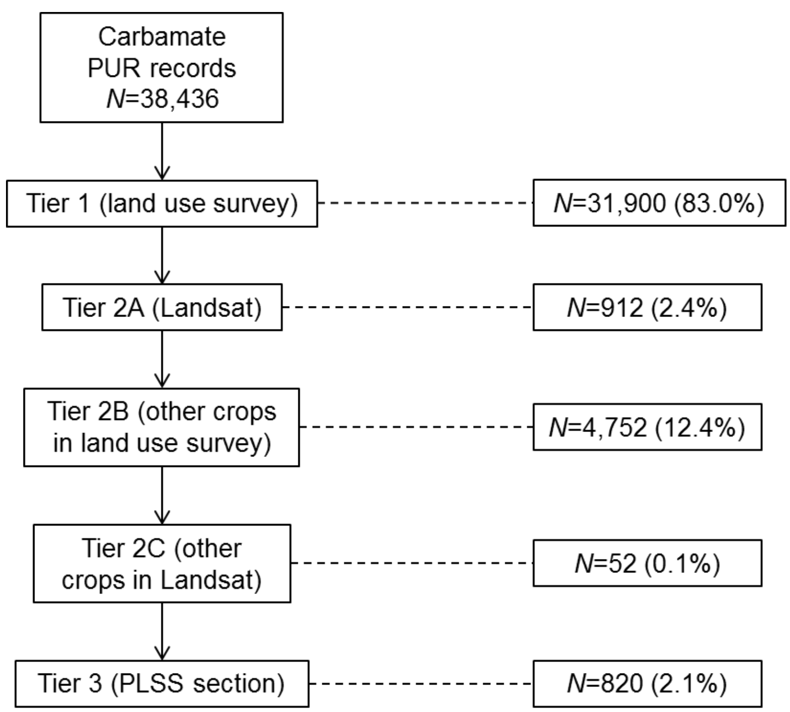


Figure 29 Carbamate PUR tier matches, Kern County (1974-1990)

4.4.1 Contribution of Landsat Imagery to Modified Three-Tier PUR Matching

According to the most specific crop-matching Tiers 1 and 2A where PUR records were matched exactly to crop type, collapsed field crops (grain and hay crops, field crops, pasture, and truck nursery and berry crops) consistently represented the majority of matches (Tables 15-17). More specifically, the Landsat-derived crop field boundaries were useful in matching collapsed field crops, in addition to the subtropical fruits (lemons, grapefruits) and deciduous fruits and nuts (walnuts, almonds, peaches and nectarines, plums) agricultural land use classes.

Table 15 Organochlorines: Tiers 1 and 2A matched crops

Tier	Crop type	*N* (%)
Tier 1 (*N*=12,355)	Collapsed field crop	12,021 (97.3%)
	Orange	261 (2.1%)
	Almond	21 (0.2%)
	Lemon	16 (0.1%)
	Plum	14 (0.1%)
	Apple	8 (0.1%)
	Peach and nectarine	6 (0.05%)
	Walnut	5 (0.04%)
	Grapefruit	3 (0.02%)
Tier 2A (*N*=304)	Collapsed field crops	270 (88.8%)
	Lemon	20 (6.6%)
	Grapefruit	9 (3.0%)
	Pear	3 (1.0%)
	Apple	1 (0.3%)
	Plum	1 (0.3%)

Table 16 Organophosphates: Tiers 1 and 2A matched crops

Tier	**Crop type**	***N* (%)**
Tier 1 (*N*=81,784)	Collapsed field crops	60,067 (73.4%)
	Almond	11,626 (14.2%)
	Orange	6,238 (7.6%)
	Peach and nectarine	1,230 (1.5%)
	Apple	781 (1.0%)
	Pistachio	614 (0.8%)
	Plum	562 (0.7%)
	Lemon	214 (0.3%)
	Walnut	183 (0.2%)
	Kiwi	76 (0.1%)
	Olive	67 (0.1%)
	Apricot	43 (0.1%)
	Grapefruit	32 (0.04%)
	Rice	21 (0.03%)
	Fig	14 (0.02%)
	Cherry	10 (0.01%)
	Prune	4 (0.005%)
	Avocado	1 (0.001%)
	Pear	1 (0.001%)
Tier 2A (*N*=2,319)	Collapsed field crops	1,539 (66.4%)
	Lemon	167 (7.2%)
	Peach and nectarine	164 (7.1%)
	Almond	146 (6.3%)
	Grapefruit	94 (4.1%)
	Plum	87 (3.8%)
	Apple	29 (1.3%)
	Kiwi	27 (1.2%)
	Orange	24 (1.0%)
	Prune	12 (0.5%)
	Walnut	10 (0.4%)
	Apricot	8 (0.3%)
	Pear	5 (0.2%)
	Miscellaneous deciduous	3 (0.1%)
	Miscellaneous subtropical fruit	2 (0.1%)
	Olive	1 (0.04%)
	Rice	1 (0.04%)

Table 17 Carbamates: Tiers 1 and 2A matched crops

Tier	Crop type	*N* (%)
Tier 1 (*N*=31,900)	Collapsed field crops	26,295 (82.4%)
	Orange	2,808 (8.8%)
	Almond	1,300 (4.1%)
	Peach and nectarine	973 (3.1%)
	Plum	143 (0.4%)
	Apple	135 (0.4%)
	Lemon	84 (0.3%)
	Pistachio	65 (0.2%)
	Miscellaneous deciduous	28 (0.1%)
	Grapefruit	23 (0.1%)
	Olive	20 (0.1%)
	Apricot	19 (0.1%)
	Cherry	7 (0.02%)
Tier 2A (*N*=912)	Collapsed field crops	573 (62.8%)
	Peach and nectarine	129 (14.1%)
	Grapefruit	58 (6.4%)
	Miscellaneous deciduous	48 (5.3%)
	Lemon	47 (5.2%)
	Almond	20 (2.2%)
	Plum	14 (1.5%)
	Apricot	8 (0.9%)
	Apple	7 (0.8%)
	Olive	3 (0.3%)
	Orange	2 (0.2%)
	Pear	2 (0.2%)
	Prune	1 (0.1%)

Figure 30 shows a specific example of a Tier 2A match, where Landsat NDVI images identified an almond field treated with organophosphate pesticides between 1974 and 1990 not otherwise found in the land use survey. An orange orchard (from land use survey) was present in section 15M27S27E06, which provided a Tier 1 match according to specific crop type and section. However, had a Tier 2A not been implemented, the

14,565.55 lb applied on an almond field would have been matched to the orange orchard using the standard Rull and Ritz (2003) approach (i.e. the 2nd tier).

Figures 31 through 33 illustrate the pounds of applied active ingredient pesticide (1974-1990) matched to any of the tiers following a spatial union of the land use survey, CSL-classified, and PLSS section layers. The spatial union overestimates the actual number of tier-matched pounds of AI due to the intersecting land use survey and Landsat crop fields. However, the union is informative in illustrating the overall pattern of pesticide applications across Kern County. As reflected in the consistently higher pesticide use with organophosphates during any given year (Figure 11), organophosphate pesticide use was more dispersed throughout Kern County compared to organochlorines and carbamates. Higher organophosphate pesticide usage was also more concentrated and more frequently occurring (≥5,000 lb) in the central and northwestern portions of the county. This geographic pattern of concentrated pesticide usage was also observed with organochlorines and carbamates, which is where a large portion of the Central Valley agricultural fields are located. Crop fields derived from land use survey, Landsat, and PLSS section data not treated with pesticides are not shown.

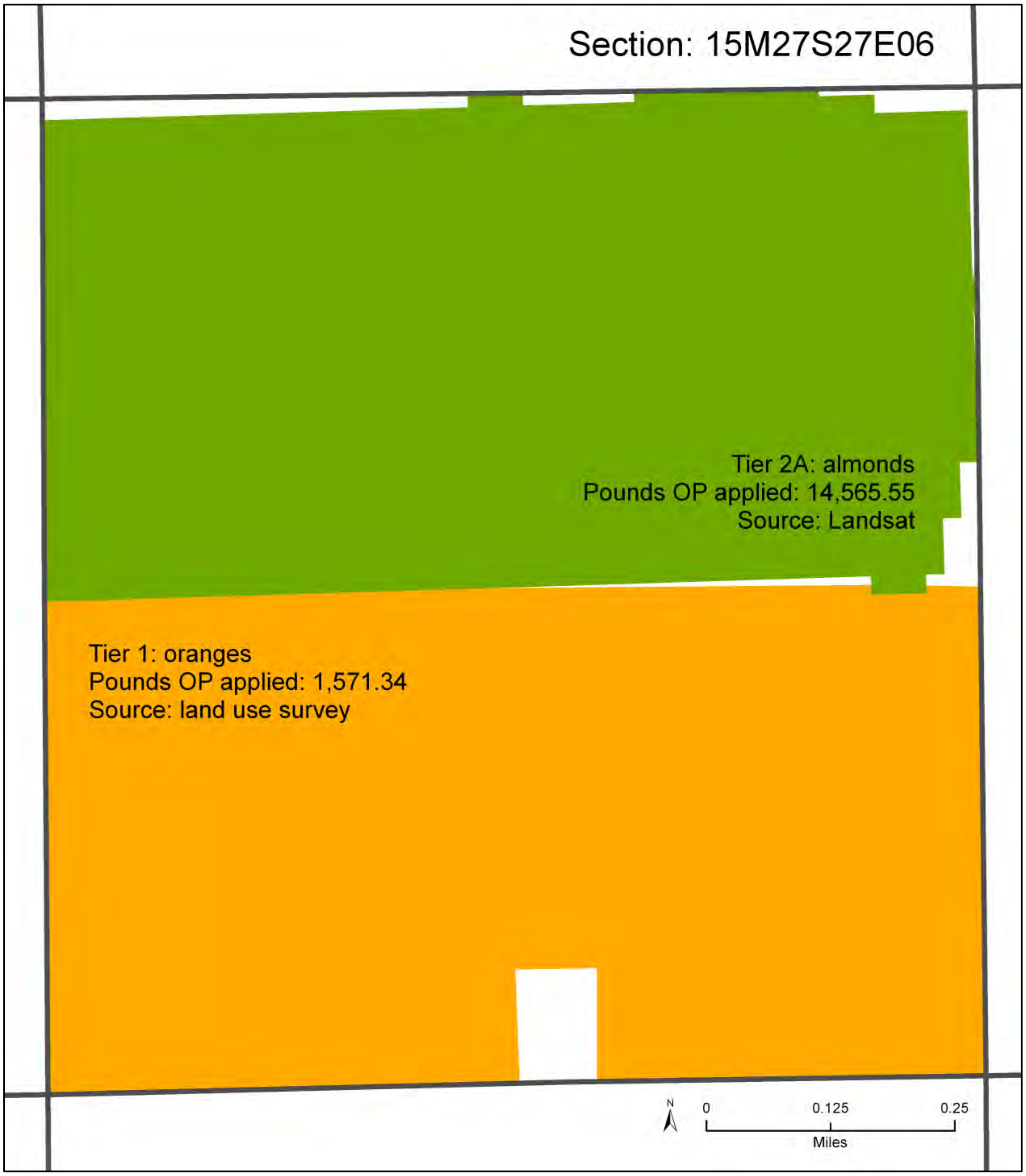


Figure 30 Tier2A match provided by Landsat, organophosphate PUR applications, 1974-1990 (Data from CDPR 2013; and CDWR 2013)

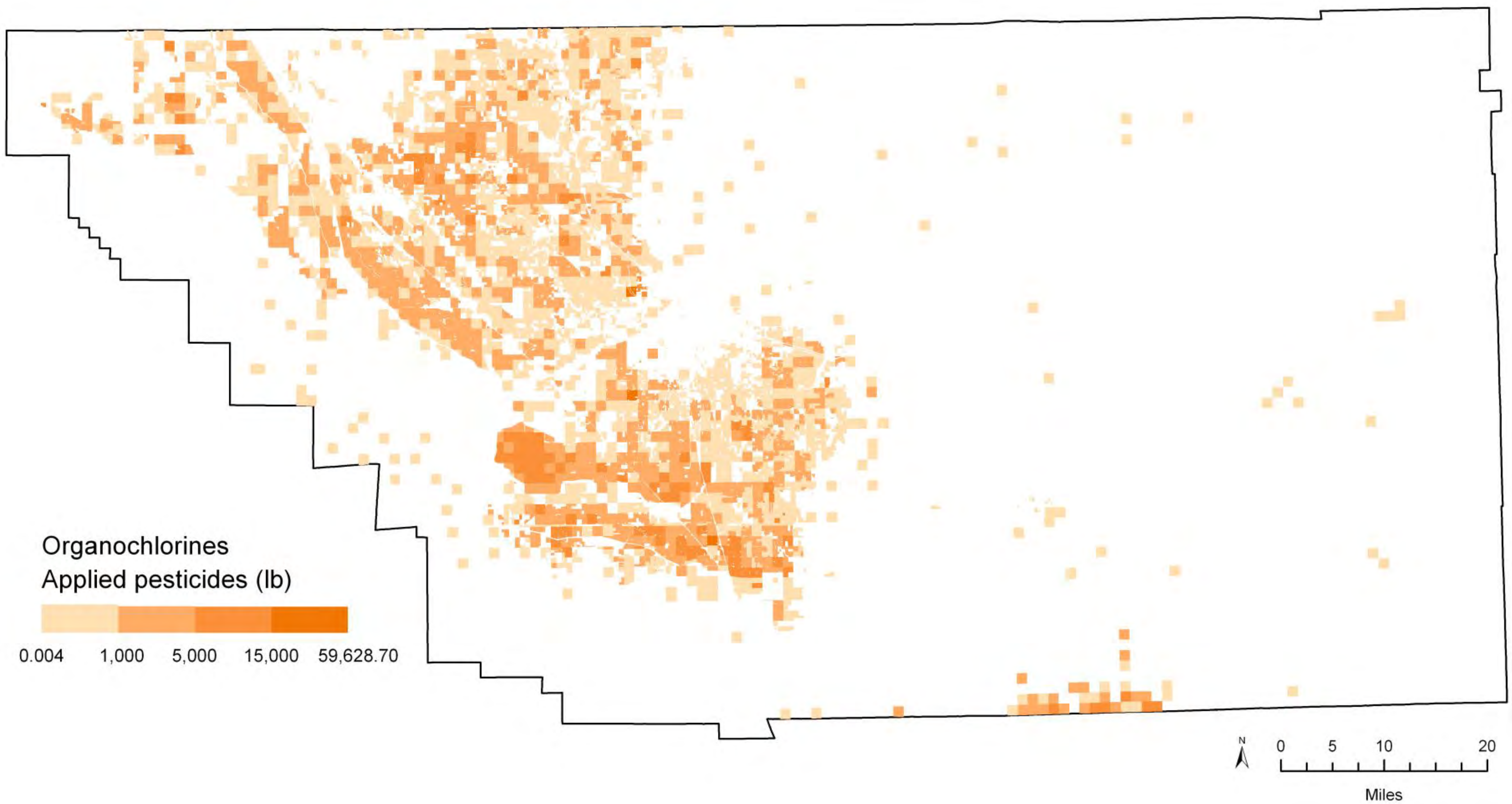


Figure 31 Organochlorines: applied pesticides on crop fields and sections, Kern County (1974-1990)
(Data from CDPR 2013; and U.S. Census Bureau 2013)

87

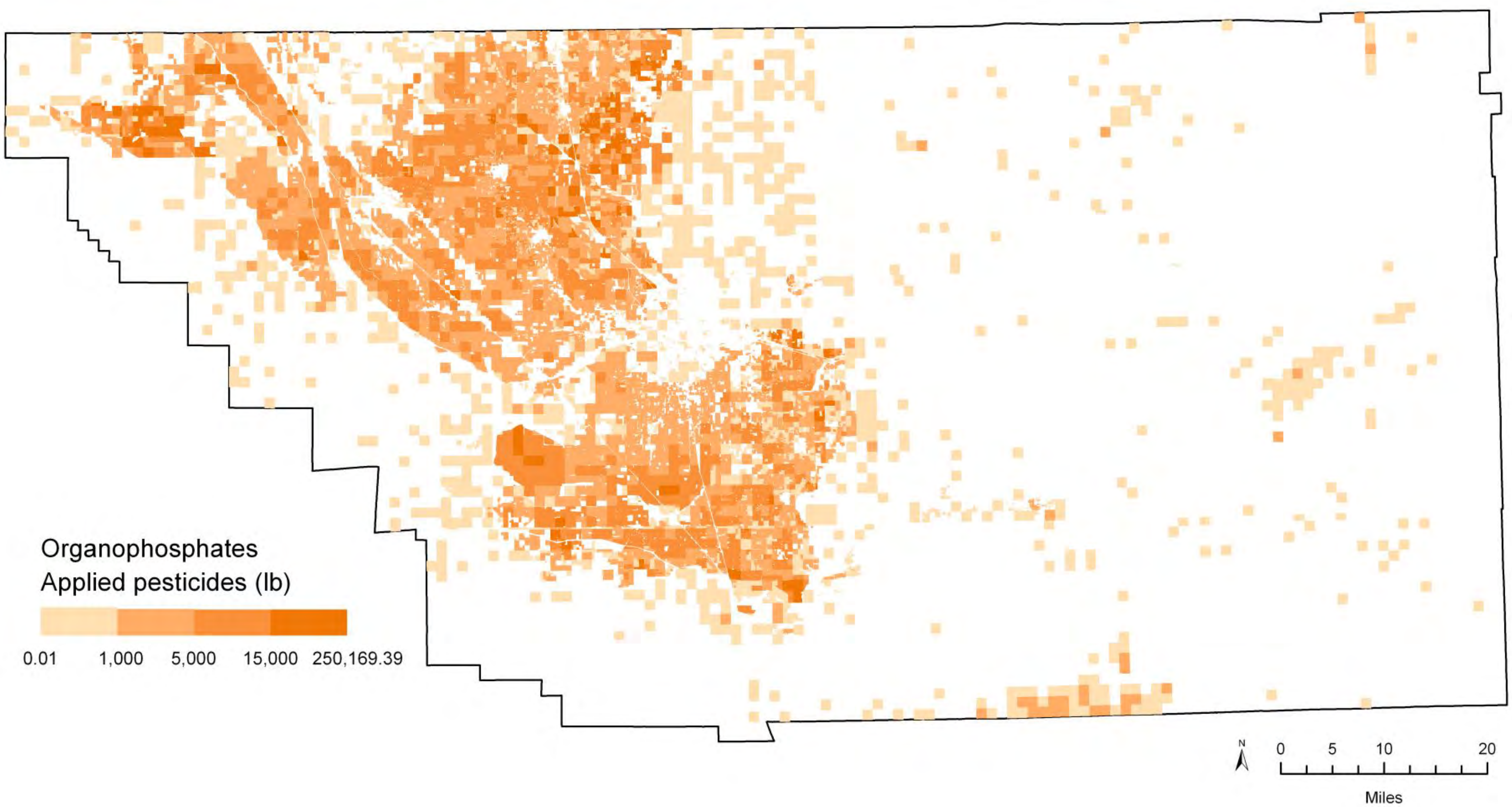


Figure 32 Organophosphates: applied pesticides on crop fields and sections, Kern County (1974-1990)
(Data from CDPR 2013; and U.S. Census Bureau 2013)

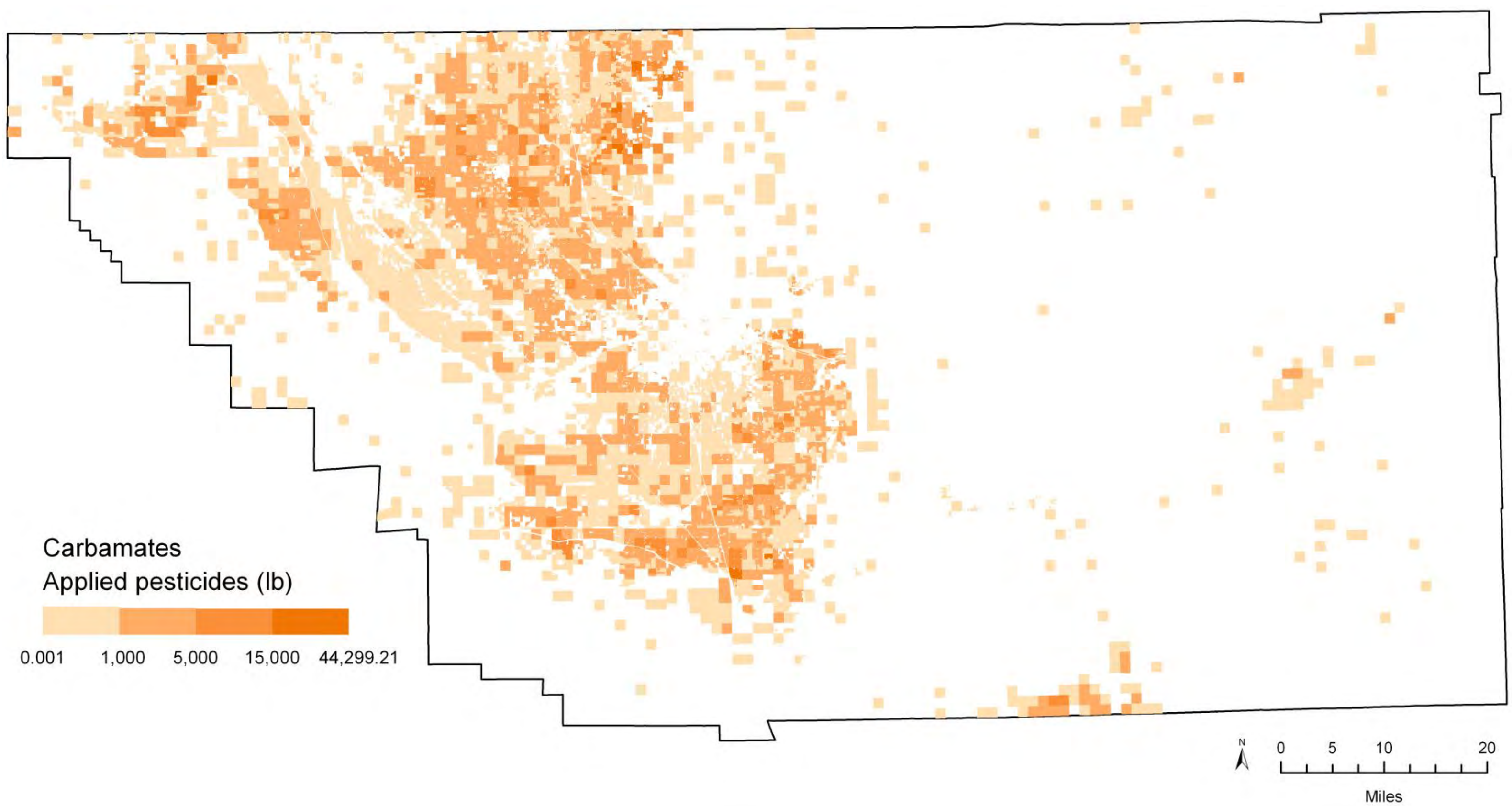


Figure 33 Carbamates: applied pesticides on crop fields and sections, Kern County (1974-1990)
(Data from CDPR 2013; and U.S. Census Bureau 2013)

89

4.5 Annual Pesticide Application Rates by Areal Aggregation

Figure 34 shows the 47 ZCTAs intersecting some portion of Kern County. The absence of ZCTAs in some areas of the county reflects the absence of ZIP codes, and thus mail delivery, in these regions (Grubesic and Matisziw 2006). At the ZCTA level, annual pesticide application rates differed according to pesticide chemical class (Figures 35-37). Consistent across all classes are higher application rates in the central and northwestern portions of Kern County. Pesticide application rates were highest for organophosphates, ranging between 0 and 1.36 lb/ac (Figure 36). This was followed by carbamates, ranging between 0 and 0.39 lb/ac (Figure 37), and organochlorines ranging between 0 and 0.25 lb/ac (Figure 35). Organochlorine usage was absent in nine ZCTAs, followed by five ZCTAs absent of carbamate usage, and two ZCTAs absent of organophosphate usage.

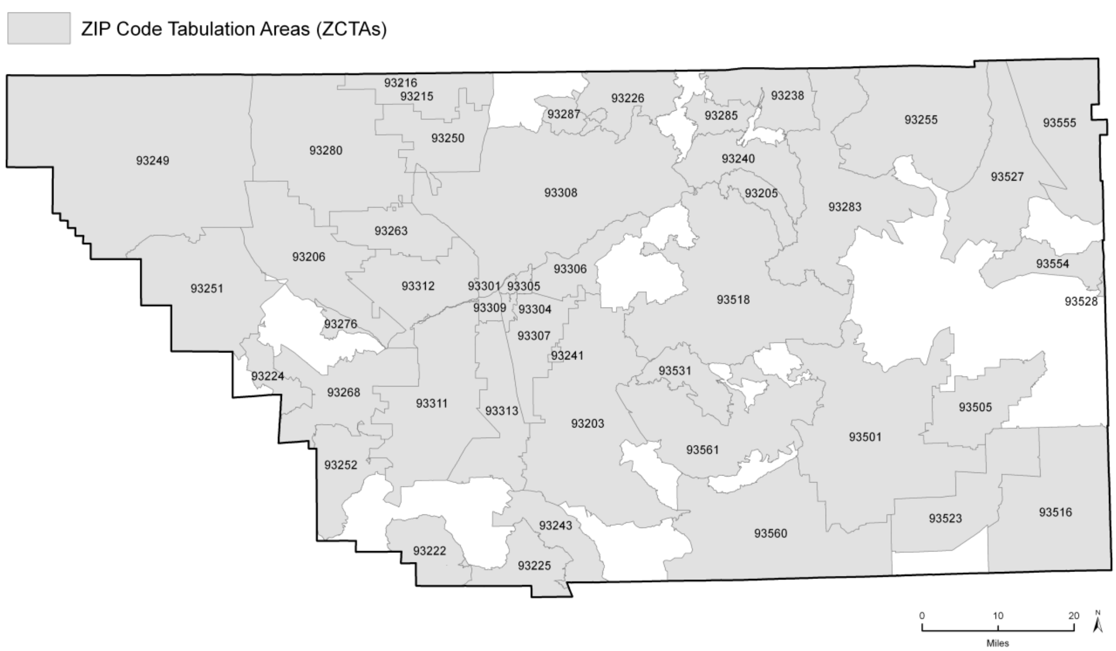


Figure 34 Kern County ZCTAs (Data from U.S. Census Bureau 2013)

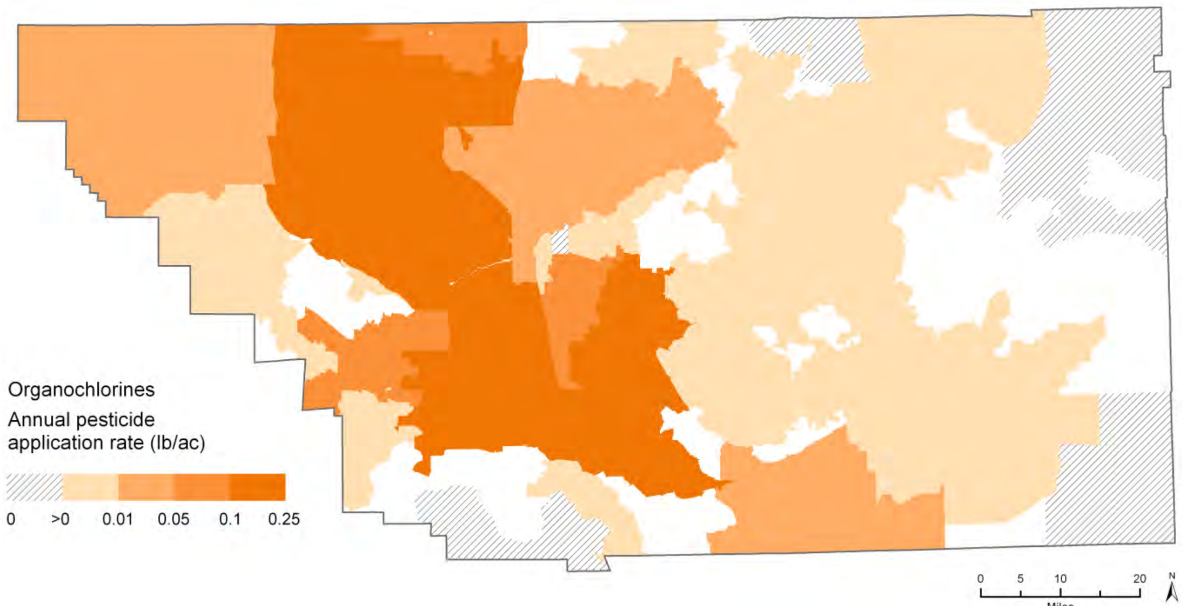


Figure 35 Organochlorines: ZCTA-level annual pesticide application rates, Kern County (1974-1990) (Data from CDPR 2013; and U.S. Census Bureau 2013)

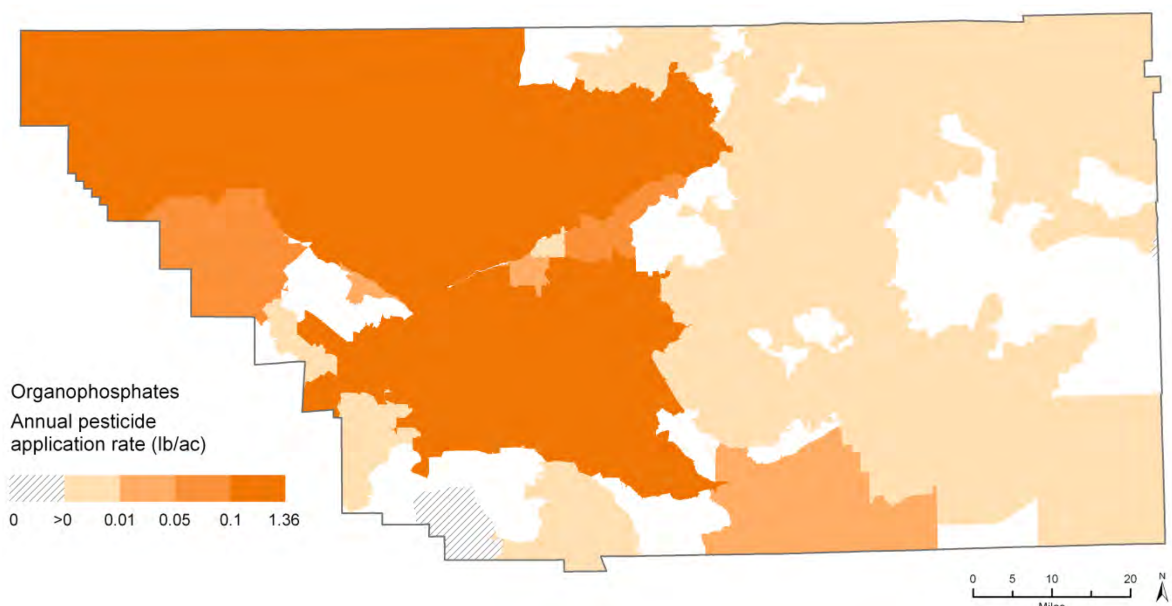


Figure 36 Organophosphates: ZCTA-level annual pesticide application rates, Kern County (1974-1990) (Data from CDPR 2013; and U.S. Census Bureau 2013)

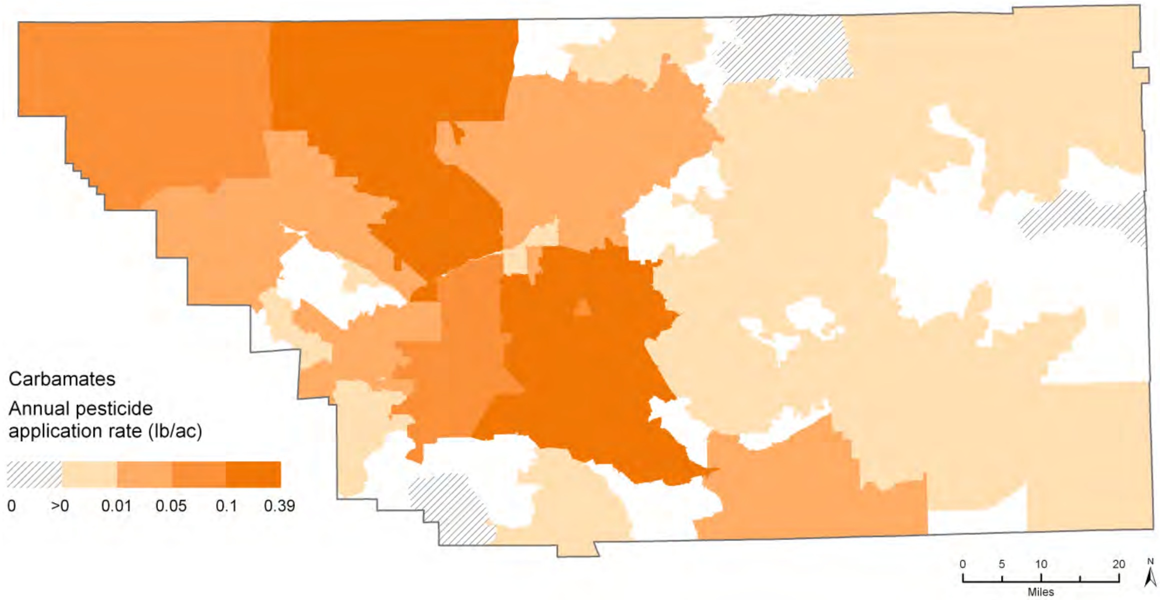


Figure 37 Carbamates: ZCTA-level annual pesticide application rates, Kern County (1974-1990) (Data from CDPR 2013; and U.S. Census Bureau 2013)

Figure 38 shows the 140 census tracts located in Kern County. Annual pesticide application rates were highest for organochlorines (maximum of 1.57 lb/ac) (Figure 39). Organophosphate usage was relatively high, upwards of 1.41 lb/ac (Figure 40). Carbamate-specific annual application rates ranged between 0 and 0.55 lb/ac (Figure 41). Looking across all three chemical classes, there was an absence of pesticide applications in central Kern County near the city of Bakersfield, but higher rates in northwestern Kern County that decreased as one moved eastward towards the Sierra Nevada Mountains. Forty census tracts were absent of organochlorine usage, followed by 29 census tracts absent of carbamate usage, and 16 census tracts absent of organophosphate usage.

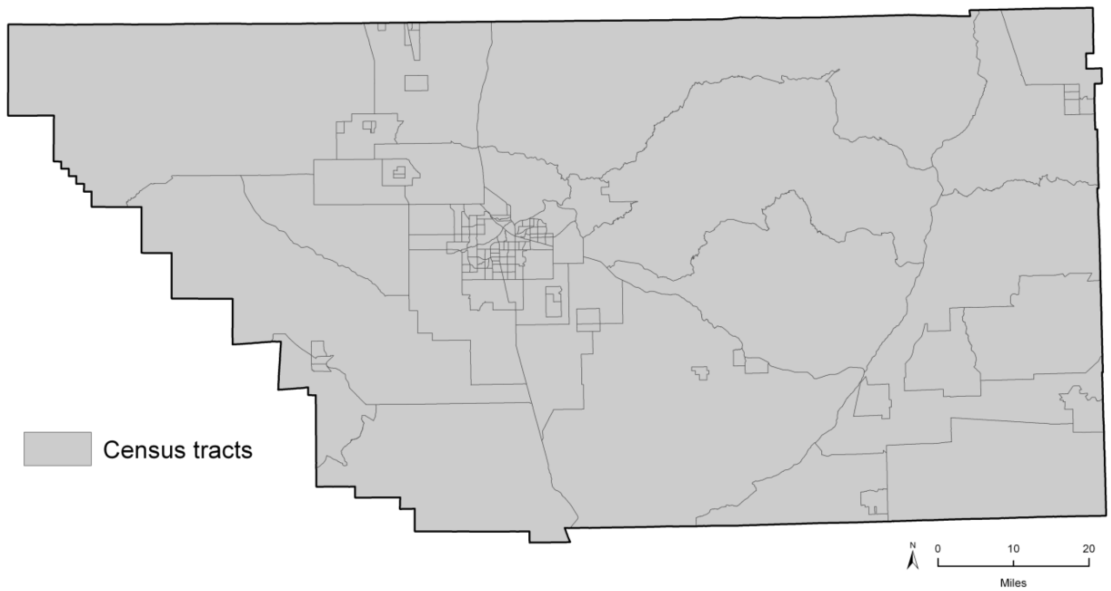


Figure 38 Kern County census tracts (Data from U.S. Census Bureau 2013)

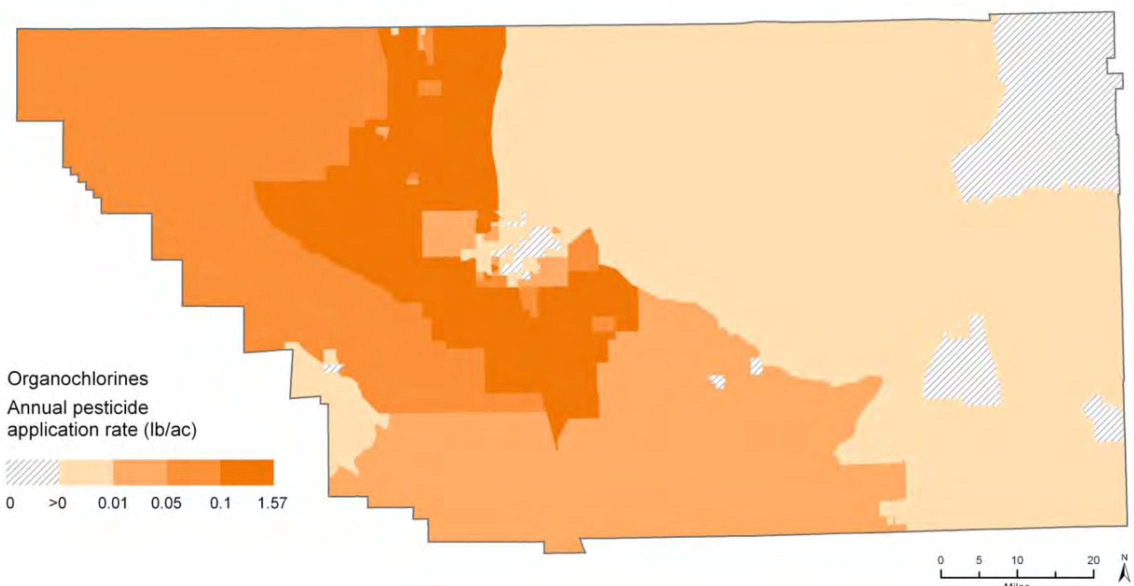


Figure 39 Organochlorines: census tract-level annual pesticide application rates, Kern County (1974-1990) (Data from CDPR 2013; and U.S. Census Bureau 2013)

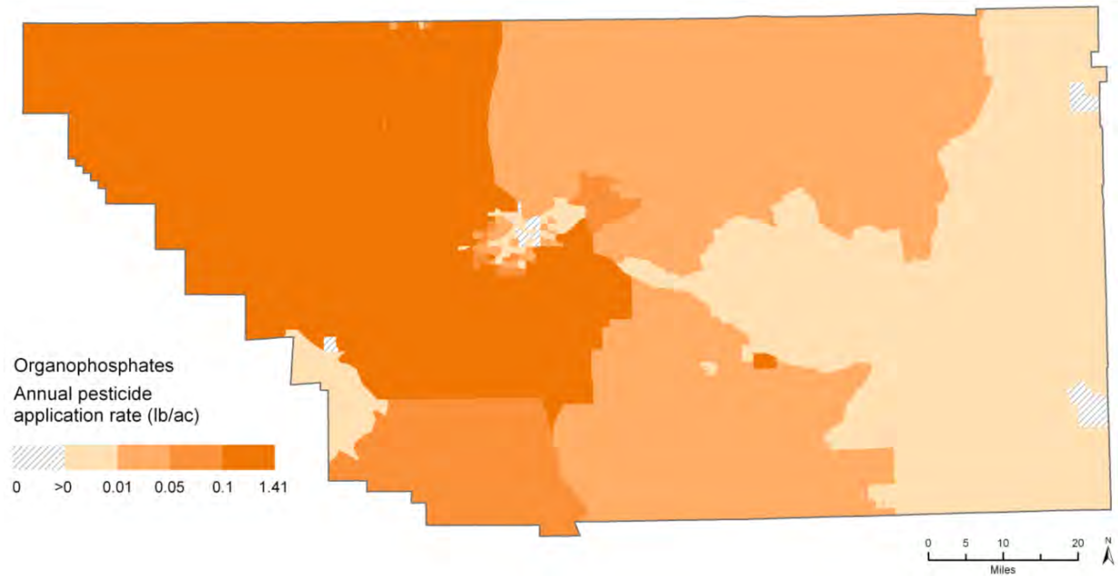


Figure 40 Organophosphates: census tract-level annual pesticide application rates, Kern County (1974-1990) (Data from CDPR 2013; and U.S. Census Bureau 2013)

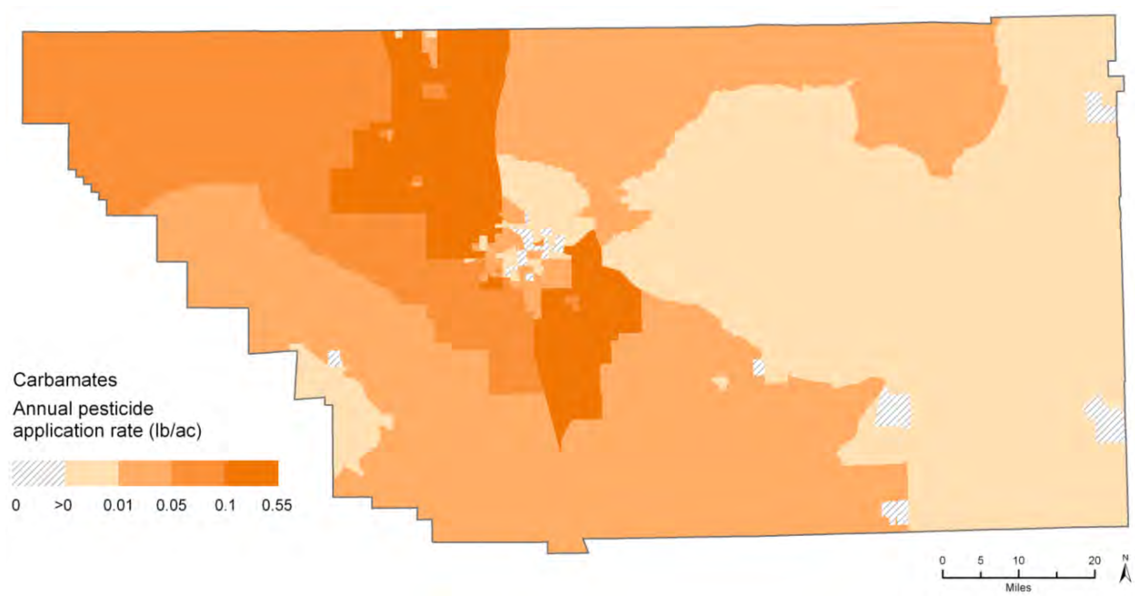


Figure 41 Carbamates: census tract-level annual pesticide application rates, Kern County (1974-1990) (Data from CDPR 2013; and U.S. Census Bureau 2013)

As a result of the choice of areal aggregations under study, some PUR tier matches were not included in calculating ZCTA-level application rates due to the absence of ZCTAs in those geographic areas. The affected PUR records matched to six land use survey crop fields and 15 PLSS sections for organochlorines, 15 land use survey crop fields, three Landsat-derived crop fields, and 80 PLSS sections for organophosphates, and 12 land use survey crop fields, two Landsat-derived crop fields, and 48 PLSS sections for carbamates. Census tract rates were not affected due to the entire extent of Kern County being covered by these areal units.

Table 18 shows the number of pesticide-treated crop fields (from land use survey or Landsat) and PLSS sections stratified by areal aggregation and pesticide chemical class. ZCTAs were more frequently intersected with organophosphate-treated crop fields and sections (median 15 fields/sections; maximum 658), which is reflected in the fewer number of ZCTAs absent of organophosphate pesticide applications (N=2). This pattern persists at the census tract level (median 4 fields/sections; maximum 828), where again, relatively few census tracts were absent of organophosphate applications (N=16). Note that a single crop field or section could have intersected multiple ZCTAs, crop fields belonging to the collapsed field crop group are represented as a single multipart polygon within a section, and that a Landsat-derived crop field may intersect a land use survey crop field.

4.6 Descriptive Analysis: Areal Aggregation and Pesticide Exposure

Despite the variation in shapes and sizes of the areal units under study, ZCTA- and census tract-level annual application rates were comparable across all three pesticide

Table 18 Pesticide-treated crop fields and sections intersecting areal units

ZCTAs (*N*=47)				
	Mean ± *SD*	**Median (*IQR*)**	**Min.**	**Max.**
Organochlorines	62.1 ± 105.8	6 (133)	0	457
Organophosphates	100.1 ± 157.7	15 (167)	0	658
Carbamates	82.4 ± 135.0	12 (145)	0	577
Census tracts (*N*=140)				
Organochlorines	22.2 ± 69.7	1 (7)	0	486
Organophosphates	36.1 ± 107.6	4 (11)	0	828
Carbamates	29.8 ± 90.2	3 (10)	0	657

chemical classes (Table 19). Census tract-level organochlorine rates were comparable to ZCTA-level rates (median 0.001 lb/ac) (p=0.5705). Census tract-level organophosphate rates were slightly higher (median 0.02 lb/ac) than ZCTA-level rates (median 0.01 lb/ac) (p=0.6104). Census tract-level carbamate rates were also comparable to ZCTA-level rates (median 0.01 lb/ac) (p=0.9801). Higher maximum application rates were typically observed when aggregated at the census tract level.

Table 19 Annual pesticide application rates according to areal aggregation

Organochlorines						
	N	**Mean ± *SD***	**Median (*IQR*)**	**Min.**	**Max.**	***p*[1]**
ZCTA	47	0.03 ± 0.06	0.001 (0.08)	0	0.25	0.5705
Census tract	140	0.04 ± 0.15	0.001 (0.03)	0	1.57	
Organophosphates						
ZCTA	47	0.16 ± 0.29	0.01 (0.30)	0	1.36	0.6104
Census tract	140	0.13 ± 0.24	0.02 (0.14)	0	1.41	
Carbamates						
ZCTA	47	0.05 ± 0.08	0.01 (0.06)	0	0.39	0.9801
Census tract	140	0.05 ± 0.09	0.01 (0.05)	0	0.55	

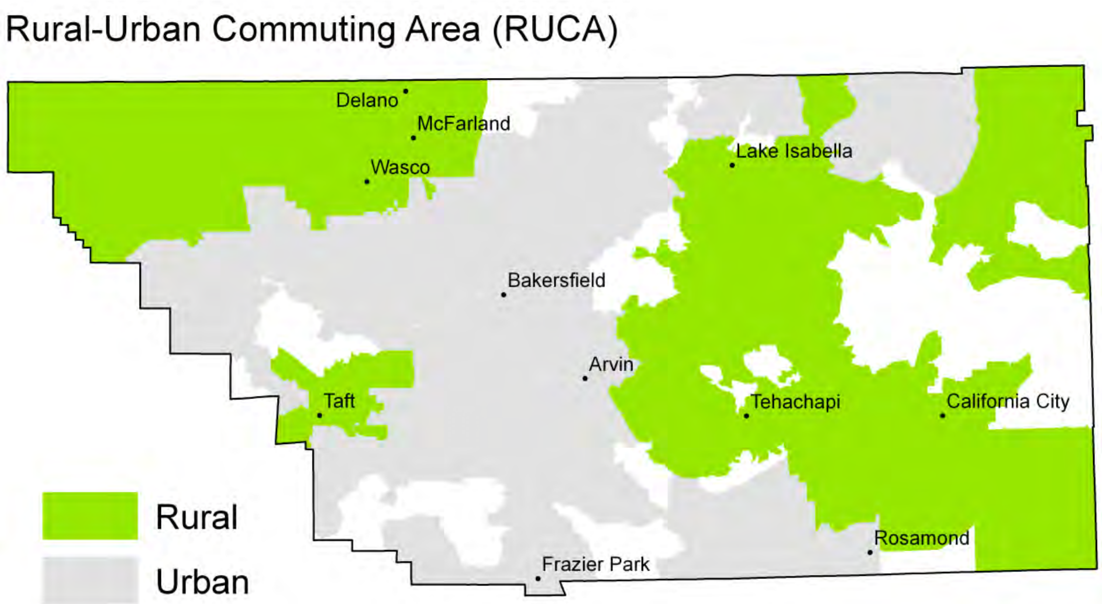
[1] Wilcoxon rank-sum test

4.7 Kern County Rurality

The geographic pattern of rurality varied according to aggregation and rurality metric. When evaluating ZCTA-level rurality, rural geographic areas were predominantly located in the western and eastern portions of Kern County for both metrics (Figure 42). Agreement in ZCTA-level rurality between RUCA codes and the U.S. Census Bureau metric was poor beyond chance (*kappa*=0.03; Table 20). Among rural ZCTAs categorized using the U.S. Census Bureau metric, 55.6% were categorized as urban using RUCA codes. Among rural ZCTAs categorized using RUCA codes, 60% were categorized as urban using the U.S. Census Bureau metric.

At the census tract level, geographic patterns of rurality were also dissimilar compared to ZCTA-level patterns. A larger proportion of Kern County was designated as rural according to RUCA codes compared to the U.S. Census Bureau metric (Figure 43). Agreement in census tract-level rurality was also poor between RUCA codes and the U.S. Census Bureau metric (*kappa*=0.04; Table 20). Among rural census tracts categorized using the U.S. Census Bureau metric, 71.4% were categorized as urban using RUCA codes. Among rural census tracts categorized using RUCA codes, 92.9% were categorized as urban using the U.S. Census Bureau metric.

Rurality designations significantly differed according to areal aggregation across both rurality metrics. When using the RUCA metric, a larger proportion of ZCTAs was categorized as rural (42.6%) compared to census tracts (20%) (p=0.0022; Table 21). When using the U.S. Census Bureau metric, a larger proportion of ZCTAs was also categorized as rural (38.3%) compared to census tracts (5%) (p<0.0001; Table 21). Given the larger number of census tracts compared to ZCTAs in Kern County, census tracts, by



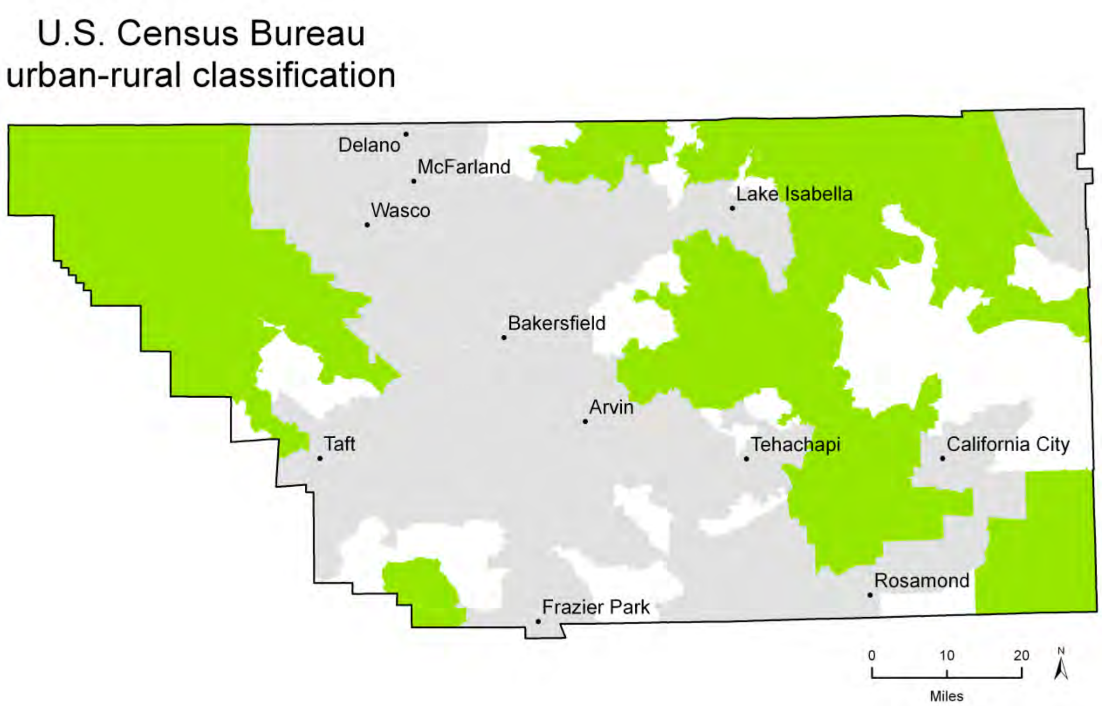


Figure 42 ZCTA-level rurality (Data from RHRC 2000; and U.S. Census Bureau 2013)

Table 20 RUCA and U.S. Census Bureau metric designations by areal aggregation

ZCTA					
		RUCA			
		Rural	Urban	Total	Kappa[1]
U.S. Census Bureau	Rural	8	10	18	0.03
	Urban	12	17	29	
	Total	20	27	47	
Census tract					
U.S. Census Bureau	Rural	2	5	7	0.04
	Urban	26	107	133	
	Total	28	112	140	

[1] Kappa values range between -1 and +1. Kappa values <0.4 indicate poor agreement beyond chance. Kappa values between 0.4 and 0.75 indicate fair to good agreement beyond chance. Kappa values >0.75 indicate excellent agreement beyond chance.

design, may be more homogeneous aggregations. In other words, since ZCTAs are relatively larger in area (median area: 56,288.28 ac), their structure may mask urban/rural differences within the ZCTA that is better captured when Kern County is partitioned according to smaller census tracts (median area: 554.5 ac), which may explain some of the differences in rurality designations by aggregation.

The extent to which rural areal units are representative of pesticide application practices is highlighted in Tables 23 and 24. Urban ZCTAs and census tracts, whether defined according to RUCA codes or the U.S. Census Bureau metric, were consistently characterized by higher application rates between 1974 and 1990 across all three pesticide chemical classes. Interestingly, rates were significantly different between urban and rural ZCTAs using the U.S. Census Bureau metric across all pesticide chemical classes. For example, carbamate application rates among urban ZCTAs (using the U.S. Census Bureau metric) were significantly higher (median 0.01 lb/ac) compared to rural ZCTAs (median 0.0005 lb/ac) (p=0.0011). Similar results were observed among census

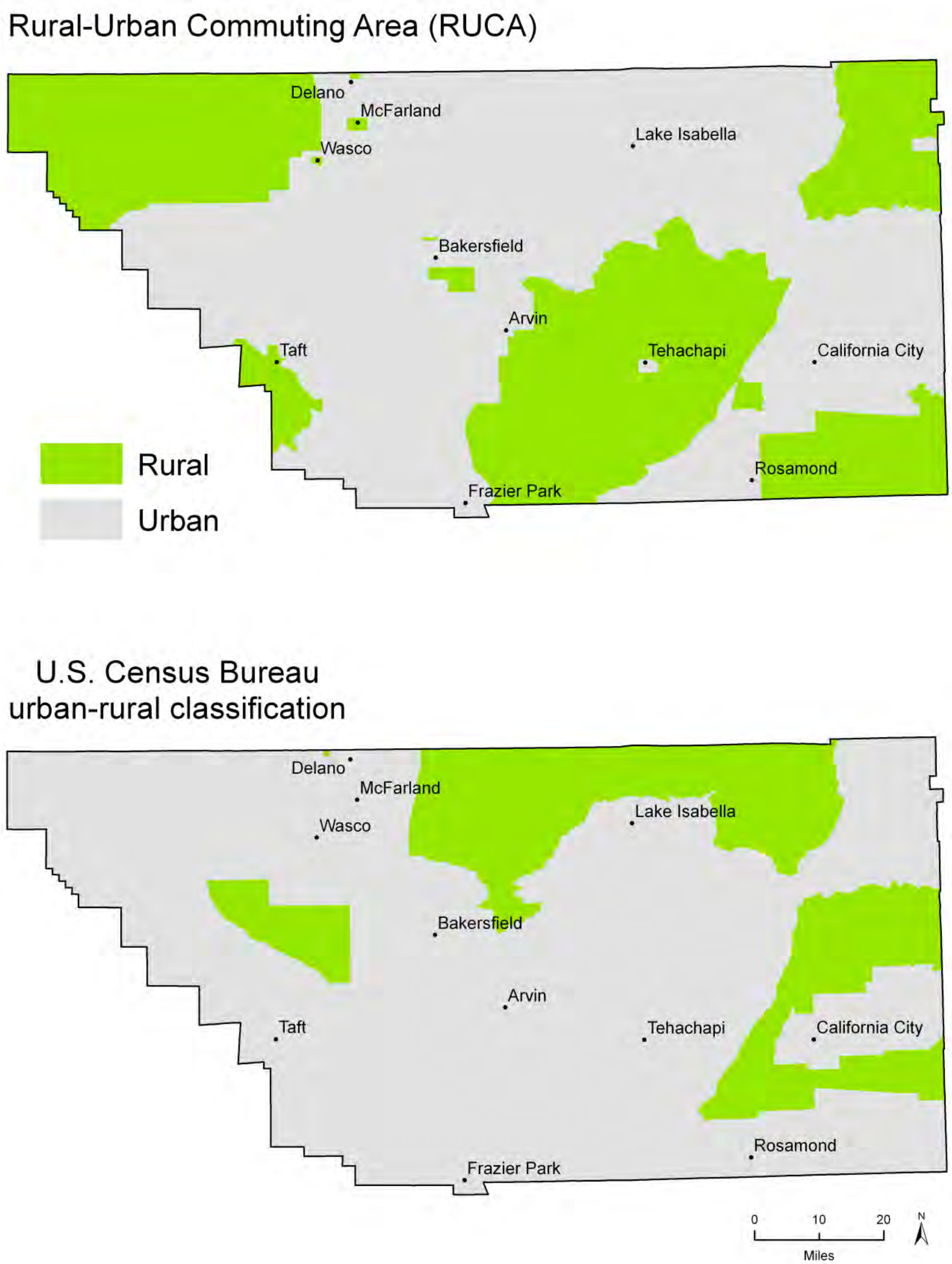


Figure 43 Census tract-level rurality.
***Sources*: Data from RHRC (2000); U.S. Census Bureau (2013).**

Table 21 ZCTA vs. census tract rurality designations

RUCA				
	Rural	**Urban**	**Total**	***p*[1,2]**
ZCTA	20	27	47	0.0022**
Census tract	28	112	140	
U.S. Census Bureau urban-rural classification				
ZCTA	18	29	47	<0.0001***
Census tract	7	133	140	

1 Chi-square test
2 ***p*<0.01; *** *p*<0.001

tracts; however, rates were significantly different according to RUCA code designations. For example, carbamate application rates among urban census tracts (using RUCA codes) were significantly higher (median 0.01 lb/ac) compared to rural census tracts (median 0.0003 lb/ac) (p=0.0074). Maximum application rates were consistently higher among urban census tracts.

The median number of pesticide-treated crop fields and sections was typically higher among urban ZCTAs (median 7-32 fields) compared to rural ZCTAs (median 1.5-10.5 fields), as reflected in the relatively higher application rates (Appendix F; Table F1). However, the median number of treated fields and sections was generally higher among rural census tracts (median 1-38 fields) vs. urban census tracts (median 1-4 fields), though the application rates do not reflect this pattern (Appendix F; Table F2). Larger urban-rural differences were observed when examining the U.S. Census Bureau metric.

Overlaying rural and urban ZCTAs and census tracts and the tiered PUR matches also mirror the results shown in Tables 22 and 23 (Appendix F; Figures F2-F7). Pesticide-treated crop fields and sections across all three chemical classes frequently intersect urban ZCTAs and census tracts along the central portions of Kern County.

Table 22 Pesticide rates stratified by rurality: ZCTAs

		N	Mean ± *SD*	Median (IQR)	Min.	Max.	$p^{1,2}$
Organochlorines							
RUCA	Rural	20	0.02 ± 0.04	0.0002 (0.020)	0	0.13	0.0533
	Urban	27	0.04 ± 0.07	0.005 (0.100)	0	0.25	
U.S. Census Bureau	Rural	18	0.01 ± 0.03	0.0001 (0.002)	0	0.11	0.0093**
	Urban	29	0.05 ± 0.07	0.005 (0.100)	0	0.25	
Organophosphates							
RUCA	Rural	20	0.20 ± 0.38	0.001 (0.22)	0	1.36	0.1822
	Urban	27	0.13 ± 0.19	0.02 (0.30)	0	0.68	
U.S. Census Bureau	Rural	18	0.04 ± 0.09	0.001 (0.01)	0	0.31	0.0056**
	Urban	29	0.24 ± 0.34	0.02 (0.37)	0.001	1.36	
Carbamates							
RUCA	Rural	20	0.05 ± 0.10	0.001 (0.040)	0	0.39	0.1186
	Urban	27	0.04 ± 0.06	0.01 (0.090)	0	0.19	
U.S. Census Bureau	Rural	18	0.01 ± 0.02	0.0005 (0.002)	0	0.06	0.0011**
	Urban	29	0.07 ± 0.10	0.01 (0.100)	0.001	0.39	

[1] Wilcoxon rank-sum test
[2] ***p*<0.01

4.7.1 Accuracy Assessment of Rurality

The accuracy of using ZCTA-level rurality metrics as a surrogate measure for pesticide exposure varied according to rurality metric, pesticide chemical class, and GIS metric (gold standard) pesticide exposure cutoff (Tables 24-29). Specificity was consistently higher than sensitivity. When evaluating RUCA codes, sensitivity was generally highest when using a 0 lb/ac cutoff, decreasing when using a 50th percentile cutoff, and

Table 23 Pesticide rates stratified by rurality: census tracts

		N	Mean ± *SD*	Median (*IQR*)	Min.	Max.	$p^{1,2}$
Organochlorines							
RUCA	Rural	28	0.01 ± 0.02	0.00004 (0.010)	0	0.10	0.0398*
	Urban	112	0.05 ± 0.17	0.003 (0.050)	0	1.57	
U.S. Census Bureau	Rural	7	0.02 ± 0.04	0.0004 (0.002)	0	0.10	0.5390
	Urban	133	0.05 ± 0.15	0.001 (0.040)	0	1.57	
Organophosphates							
RUCA	Rural	28	0.05 ± 0.12	0.003 (0.05)	0	0.55	0.0078**
	Urban	112	0.15 ± 0.25	0.03 (0.19)	0	1.41	
U.S. Census Bureau	Rural	7	0.06 ± 0.12	0.01 (0.06)	0	0.32	0.5822
	Urban	133	0.14 ± 0.24	0.02 (0.14)	0	1.41	
Carbamates							
RUCA	Rural	28	0.01 ± 0.03	0.0003 (0.01)	0	0.09	0.0074**
	Urban	112	0.06 ± 0.10	0.01 (0.08)	0	0.55	
U.S. Census Bureau	Rural	7	0.01 ± 0.02	0.003 (0.02)	0	0.06	0.4538
	Urban	133	0.05 ± 0.10	0.01 (0.05)	0	0.55	

1 Wilcoxon rank-sum test
2 *p<0.05; **p<0.01

increasing when using a 75th percentile cutoff. Specificity when using RUCA codes followed a similar pattern, where specificity decreased when using a 50th percentile cutoff compared to a 0 lb/ac cutoff, but was highest when using a 75th percentile cutoff. On the other hand, when examining the U.S. Census Bureau metric, sensitivity decreased and specificity increased as the pesticide exposure cutoffs became more conservative.

Sensitivity ranged between 25 and 42.9% for RUCA codes. In other words, the probability of a ZCTA being classified as rural given the ZCTA was pesticide-exposed ranged between 0.25 and 0.429 - where a probability of 1 is perfect sensitivity. Another to express this result is to state that RUCA codes correctly identified between 25 and

Table 24 ZCTA-level accuracy of RUCA codes: organochlorines

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	$p^{1,2}$
RUCA	Exposed	15	5	39.5%	44.4%	0.4653
	Not exposed	23	4			
Pesticide exposure cutoff: ≥50th percentile (0.001 lb/ac)						
RUCA	Exposed	6	14	25.0%	39.1%	0.0129*
	Not exposed	18	9			
Pesticide exposure cutoff: ≥75th percentile (0.08 lb/ac)						
RUCA	Exposed	4	16	33.3%	54.3%	0.4541
	Not exposed	8	19			

1 Chi-square test or Fisher's exact test
2 *p<0.05

Table 25 ZCTA-level accuracy of U.S. Census Bureau urban-rural classification: organochlorines

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	$p^{1,2}$
U.S. Census Bureau	Exposed	12	6	31.6%	33.3%	0.0676
	Not exposed	26	3			
Pesticide exposure cutoff: ≥50th percentile (0.001 lb/ac)						
U.S. Census Bureau	Exposed	5	13	20.8%	43.5%	0.0119*
	Not exposed	19	10			
Pesticide exposure cutoff: ≥75th percentile (0.08 lb/ac)						
U.S. Census Bureau	Exposed	1	17	8.3%	51.4%	0.0167*
	Not exposed	11	18			

1 Chi-square test or Fisher's exact test
2 *p<0.05

Table 26 ZCTA-Level accuracy of RUCA codes: organophosphates

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	p[1]
RUCA	Exposed	19	1	42.2%	50.0%	>0.99
	Not exposed	26	1			
Pesticide exposure cutoff: ≥50th percentile (0.01 lb/ac)						
RUCA	Exposed	7	13	29.2%	43.5%	0.0579
	Not exposed	17	10			
Pesticide exposure cutoff: ≥75th percentile (0.30 lb/ac)						
RUCA	Exposed	4	16	33.3%	54.3%	0.4541
	Not exposed	8	19			

1 Chi-square test or Fisher's exact test

Table 27 ZCTA-level accuracy of U.S. Census Bureau urban-rural classification: organophosphates

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	p[1,2]
U.S. Census Bureau	Exposed	16	2	35.6%	0%	0.1415
	Not exposed	29	0			
Pesticide exposure cutoff: ≥50th percentile (0.01 lb/ac)						
U.S. Census Bureau	Exposed	5	13	20.8%	43.5%	0.0119*
	Not exposed	19	10			
Pesticide exposure cutoff: ≥75th percentile (0.30 lb/ac)						
U.S. Census Bureau	Exposed	1	17	8.3%	51.4%	0.0167*
	Not exposed	11	18			

1 Chi-square test or Fisher's exact test
2 *p<0.05

Table 28 ZCTA-level accuracy of RUCA codes: carbamates

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	$p^{1,2}$
RUCA	Exposed	18	2	42.9%	60.0%	>0.99
	Not exposed	24	3			
Pesticide exposure cutoff: ≥50th percentile (0.01 lb/ac)						
RUCA	Exposed	6	14	26.1%	41.7%	0.0254*
	Not exposed	17	10			
Pesticide exposure cutoff: ≥75th percentile (0.06 lb/ac)						
RUCA	Exposed	5	15	41.7%	57.1%	0.9426
	Not exposed	7	20			

1 Chi-square test or Fisher's exact test
2 *p<0.05

Table 29 ZCTA-level accuracy of U.S. Census Bureau urban-rural classification: carbamates

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	$p^{1,2}$
U.S. Census Bureau	Exposed	13	5	31.0%	0%	0.0056**
	Not exposed	29	0			
Pesticide exposure cutoff: ≥50th percentile (0.01 lb/ac)						
U.S. Census Bureau	Exposed	3	15	13.0%	37.5%	0.0005**
	Not exposed	20	9			
Pesticide exposure cutoff: ≥75th percentile (0.06 lb/ac)						
U.S. Census Bureau	Exposed	1	17	8.3%	51.4%	0.0167*
	Not exposed	11	18			

1 Chi-square test or Fisher's exact test
2 *p<0.05; **p<0.01

way 42.9% of all truly pesticide-exposed ZCTAs. The remaining 57.1 to 75% of ZCTAs represent false negatives, or ZCTAs that were incorrectly classified as urban. Sensitivity ranged between 8.3 and 35.6% for the U.S. Census Bureau metric.

Specificity ranged between 39.1 and 60% for RUCA codes. In other words, between 39.1 and 60% of all truly unexposed ZCTAs were classified as urban. The remaining 40 to 60.9% of ZCTAs represent false positives, or ZCTAs incorrectly classified as rural. Specificity ranged between 0 and 51.4% for the U.S. Census Bureau metric.

Significant differences were observed when comparing the GIS gold standard to the U.S. Census Bureau metric - depending on chemical class - across all pesticide exposure cutoffs. For example, a larger proportion of pesticide-exposed ZCTAs (50th percentile) were false negatives compared to true positives. In other words, a substantial proportion of pesticide-exposed ZCTAs were misclassified as urban. Eighty-seven percent of carbamate-exposed ZCTAs were false negatives (using 50th percentile), while 13% were true positives (p=0.0005; Table 29). Fewer significant differences were observed when comparing RUCA codes to the GIS gold standard.

The accuracy of census tract-level rurality metrics also differed according to rurality metric, pesticide chemical class, and pesticide exposure cutoff. Specificity was consistently high, upwards of 77.1% for RUCA codes, and upwards of 97.5% for the U.S. Census Bureau metric (Tables 30-35). In other words, RUCA codes classified at most 77.1% of census tracts not exposed to pesticides as urban, and the U.S. Census Bureau metric classified at most 97.5% of census tracts not exposed to pesticides as urban.

Table 30 Census tract-level accuracy of RUCA codes: organochlorines

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	p[1]
RUCA	Exposed	17	11	17.0%	72.5%	0.1606
	Not exposed	83	29			
Pesticide exposure cutoff: ≥50th percentile (0.001 lb/ac)						
RUCA	Exposed	10	18	14.3%	74.3%	0.0910
	Not exposed	60	52			
Pesticide exposure cutoff: ≥75th percentile (0.03 lb/ac)						
RUCA	Exposed	4	24	11.4%	77.1%	0.1432
	Not exposed	31	81			

[1] Chi-square test

Table 31 Census tract-level accuracy of U.S. Census Bureau urban-rural classification: organochlorines

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	p[1]
U.S. Census Bureau	Exposed	6	1	6.0%	97.5%	0.6730
	Not exposed	94	39			
Pesticide exposure cutoff: ≥50th percentile (0.001 lb/ac)						
U.S. Census Bureau	Exposed	2	5	2.9%	92.9%	0.4411
	Not exposed	68	65			
Pesticide exposure cutoff: ≥75th percentile (0.03 lb/ac)						
U.S. Census Bureau	Exposed	1	6	2.9%	94.3%	0.6801
	Not exposed	34	99			

[1] Fisher's exact test

Table 32 Census tract-level accuracy of RUCA codes: organophosphates

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	*p*[1,2]
RUCA	Exposed	24	4	19.4%	75.0%	0.5269
	Not exposed	100	12			
Pesticide exposure cutoff: ≥50th percentile (0.02 lb/ac)						
RUCA	Exposed	8	20	11.4%	71.4%	0.0112*
	Not exposed	62	50			
Pesticide exposure cutoff: ≥75th percentile (0.14 lb/ac)						
RUCA	Exposed	2	26	5.7%	75.2%	0.0147*
	Not exposed	33	79			

1 Chi-square test or Fisher's exact test
2 *p<0.05

Table 33 Census tract-level accuracy of U.S. Census Bureau urban-rural classification: organophosphates

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	*p*[1]
U.S. Census Bureau	Exposed	6	1	4.8%	93.8%	0.5809
	Not exposed	118	15			
Pesticide exposure cutoff: ≥50th percentile (0.02 lb/ac)						
U.S. Census Bureau	Exposed	3	4	4.3%	94.3%	>0.99
	Not exposed	67	66			
Pesticide exposure cutoff: ≥75th percentile (0.14 lb/ac)						
U.S. Census Bureau	Exposed	1	6	2.9%	94.3%	0.6801
	Not exposed	34	99			

1 Fisher's exact test

Table 34 Census tract-level accuracy of RUCA codes: carbamates

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	p^1
RUCA	Exposed	21	7	18.9%	75.9%	0.5316
	Not exposed	90	22			
Pesticide exposure cutoff: ≥50th percentile (0.01 lb/ac)						
RUCA	Exposed	8	20	11.4%	71.4%	0.0112*
	Not exposed	62	50			
Pesticide exposure cutoff: ≥75th percentile (0.05 lb/ac)						
RUCA	Exposed	3	25	8.6%	76.2%	0.0510
	Not exposed	32	80			

1 Chi-square test
2 *p<0.05

Table 35 Census tract-level accuracy of U.S. Census Bureau urban-rural classification: carbamates

Pesticide exposure cutoff: >0 lb/ac						
		GIS metric (gold standard)				
		Exposed	Not exposed	Sensitivity	Specificity	p^1
U.S. Census Bureau	Exposed	5	2	4.5%	93.1%	0.6344
	Not exposed	106	27			
Pesticide exposure cutoff: ≥50th percentile (0.01 lb/ac)						
U.S. Census Bureau	Exposed	3	4	4.3%	94.3%	>0.99
	Not exposed	67	66			
Pesticide exposure cutoff: ≥75th percentile (0.05 lb/ac)						
U.S. Census Bureau	Exposed	1	6	2.9%	94.3%	0.6801
	Not exposed	34	99			

1 Chi-square test

Sensitivity was relatively low across both rurality metrics and all pesticide chemical classes, ranging from 5.7 to 19.4% for RUCA codes and from 2.9 to 6% for the U.S. Census Bureau metric. When examining RUCA codes, sensitivity decreased when using more conservative pesticide exposure cutoffs, while specificity remained relatively constant across all pesticide exposure cutoffs. When examining the U.S. Census Bureau metric, sensitivity and specificity remained constant across pesticide exposure cutoffs.

Significant differences were observed when comparing RUCA codes to the GIS gold standard. Among organophosphate and carbamate usage, a large proportion of pesticide-exposed census tracts (using the 50th or 75th percentile) were misclassified as urban. For example, 5.7% of organophosphate-exposed census tracts (75th percentile) were classified as rural and 94.3% were misclassified as urban (p=0.0147; Table 32). Similar results were observed for carbamates using a 50th percentile cutoff (p=0.0112; Table 34).

CHAPTER FIVE: DISCUSSION AND CONCLUSIONS

Landsat satellite-borne imagery represents an invaluable resource (data available beginning in 1972) that can be integrated into pesticide exposure methodologies (USGS 2013b). Given its moderate spatial and temporal resolution - the Landsat Thematic Mapper sensors acquired images with at least 30 m spatial resolution every 16 days - Landsat has the capacity to contribute information in determining likely crop field locations of PUR pesticide applications occurring in the past. This research demonstrated the feasibility of incorporating Landsat imagery into a GIS-based method to improve the spatiotemporal resolution of pesticide exposure estimation, representing a large-scale implementation of the Maxwell (2011) Landsat methods across Kern County (Maxwell et al. 2010b; Maxwell 2011). The presented pesticide exposure approach modified an existing individual-level, GIS-based three-tier method (Rull and Ritz 2003) that incorporates pesticide, land use, and cadastral datasets to accommodate the ZIP code and census tract analysis scales and to integrate Landsat remote sensing. This modified three-tier approach can be adopted for analysis scales finer than aggregated areal units.

Furthermore, comparing the validity of two commonly used rurality metrics, RUCA codes and the U.S. Census Bureau urban-rural classification system, provided clarity with respect to which measure is a superior surrogate of pesticide exposure to use in terms of accuracy. These rurality metrics may prove valuable in situations (i.e. places) with limited pesticide information. Hence, the absence of pesticide information, lack of technical knowledge, etc., may require the use of rurality-based metrics. The results highlight wide variability in terms of sensitivity and specificity as a function of rurality metric, areal aggregation, pesticide chemical class, and pesticide exposure cutoff.

5.1 Critical Assessment of Methods and Results: Strengths and Limitations

5.1.1 PUR Processing

The pesticide database of organochlorines, organophosphates, and carbamates included 157 pesticide active ingredients (AIs). The pesticide database may not have included all possible pesticide AIs in each chemical class, potentially underestimating pounds of pesticide used, and thus ZCTA- and census tract-level annual application rates. Regarding PUR data processing, although applying logic checks is a conservative PUR data cleaning approach that has not been documented in the literature, these methods were used by the CDPR to evaluate PUR data quality (CDPR 2000a). Applying two of the three 1990-onward PUR outlier definitions to PUR data between 1974 and 1989 has also never been performed in previous research, but adds consistency to PUR handling.

5.1.2 Crop Signature Library (CSL)

The CSL was a major component of this research, spanning January to October 1990 and including NDVI values for sampled land use survey polygons acquired via stratified random sampling (SRS). The CSL formed the basis for the classification of 1985 Landsat imagery to be incorporated into the modified three-tier pesticide exposure methodology. Its major strength and improvement upon previous CSL-related research was harnessing all available spectral information in the form of NDVI values from each SRS-sampled land use survey polygon. This represented an objective alternative to the Maxwell (2010) approach, which selects one pixel per polygon at the location of the polygon label - except when the spectral tone of the pixel is not representative of the polygon.

A potential limitation relates to the mixed pixel problem, or when a pixel is not completely occupied by a single homogeneous category (Campbell and Wynne 2011). This is a common issue at the edges of large, discrete objects and linear features. A scene divided into discrete pixel areas averages the brightness values over the entire pixel area. As the geographic features under study (crop fields; mean area of SRS polygons 407,928.4 m^2, median area 176,249.8 m^2) are larger than the pixels of the Landsat images (30 m^2), the CSL represents an H-resolution model, characterized by spectral responses of features mixed together so that composite signatures of the pixels do not match the pure spectral signature(s) (Strahler et al. 1986). For example, as NDVI values for each SRS-sampled land use survey polygon were extracted, NDVI values of pixels along the edges of the polygons may represent NDVI values of multiple crop types by virtue of the spatial resolution of the sensor. However, this issue may not have been impactful regarding the CSL if neighboring land use survey polygons were relatively uniform with respect to NDVI values, despite being associated with different crop types.

Obtaining high quality, cloud-free Landsat images was challenging. Cloud cover affected the inclusion of entire Path-Row images for March and June 1990. Furthermore, Landsat images were not available for November and December 1990, which would have contributed to a more complete CSL and enhanced the discrimination of 1985 Landsat images that were subsequently used to create an agricultural crop field layer. In addition, including multiple images within the same month, rather than limiting the CSL to one image per month, would have better captured intra-month NDVI variability and further improved classification.

Another prominent issue was the shrinking geographic coverage from which a CSL could be created. Affected by cloud cover and Path-Row data availability, a modestly sized region of northwestern Kern County was used to execute SRS (Figure 19). Although SRS ensured that all land use classes were sampled irrespective of their prevalence across Kern County, the potential for random error was introduced as 49 land use classes included SRS samples less than the *a priori* specified stratum size of 30. SRS samples within these 49 strata may not be representative of the land use class, and may have been characterized by atypical NDVI values. Expansion of the study area beyond Kern County would increase the population of each land use class strata from which polygons could be sampled.

Cloud cover and shadows were addressed at three stages of CSL creation and 1985 Landsat image classification: (1) selection of images for inclusion in the analysis; (2) masking cloud-affected areas before mosaicking Path-Row images together; and (3) digitizing geographic areas without cloud cover before SRS or PCA. Note that if clouds were present in an image but were not within a region overlapping Kern County, these areas were not masked out but were eventually excluded after cropping each radiometrically and atmospherically processed image to the Kern County extent. Some Landsat images may have been salvageable, with geographic areas covering some geographic portions of Kern County. However, cloud cover-related decisions were ultimately guided by the amount of cloud-free areas intersecting Kern County and anticipated mosaicking difficulties.

The major limiting factor affecting the CSL was the 1990 Kern County land use survey used as the ground truth. The CSL is only as informative as its source data - the land use survey guided the extraction of NDVI values for land use classes according to land use survey-delineated crop field boundaries. Land use surveys focus on agricultural land use (CDWR 2013). Therefore, other land use classes related to non-agricultural uses may not be adequately captured in this dataset. This issue is manifest in the large number of segments classified as jojoba - nine land use survey crop fields in 1990 vs. 1,572 segments were classified as jojoba with Landsat imagery. The external validity, or generalizability, of the CSL is limited by the study area of Kern County and the eligibility criteria applied prior to selecting SRS samples: single-use crop fields, at least 4 ha in area, and within the designated cloud-free zone. Grain and hay crops were the most frequently occurring multi-use or inter-cropped land use class; their exclusion from the CSL may have limited the ability of the CSL to classify segments. Results may have been impacted by sources of positional and attribute error in the data sources. For example, data entry error may have misclassified land use survey crop field names. The positional accuracy of the crop fields is also affected by methods from which boundaries were delineated (e.g. GPS). Furthermore, the extent to which the CSL is consistently representative of monthly NDVI spectral profiles for the included crop types during time periods before and after 1990 should be explored.

5.1.3 Classification of 1985 Landsat Imagery

Landsat images from 1985 selected for CSL-based classification were also affected by cloud cover for the months of January, March, and October. Particular Path-Row scenes were altogether missing, limiting the geographic extent of classified crop fields. Ideally, all Path-Row scenes would have been available for all months in 1985 paralleling the

1990 CSL (January to October). However, a decision was made to assign more weight to the contribution of more time points with NDVI values with respect to classification rather than capturing as much geographic extent as possible. In other words, only geographic areas with NDVI image data availability for all months (January, March to October; February was excluded due to absence of Path 42) were selected for subsequent PCA and segmentation. An alternative would have been to include as much of Kern County as possible by choosing only to consider months where all Path-Row scenes were available (April-June, September), which would have produced different classified crop fields. Although the alternative strategy would have resulted in a classified crop field layer covering all of Kern County, it may not have been able to adequately discriminate between land use classes, as only including four months of NDVI data may have misclassified land use classes with similar spectral profiles during these months.

A key component of this research that should be addressed is the implementation of a formal accuracy assessment of the classified Landsat imagery. In practice, classified remote sensing products should be compared to a gold standard to derive an error matrix (Campbell and Wynne 2011). Measures such as user's and producer's accuracy, as well as kappa, can be calculated to quantify the extent to which the classified product is a valid representation of the phenomenon it seeks to represent. Future research should explore the validity of this CSL in terms of classifying crop fields in Kern County at different points in time and also in different geographic areas around California.

5.1.4 Segmentation

The segmentation process is subjective in allowing the end user to specify particular parameters (e.g. tolerance in IDRISI Selva) to achieve a segmented layer of objects representing a satisfactory likeness of the geographic phenomenon under study - crop field boundaries in this analysis. A further challenge related to the need to account for differences in crop field boundaries over the course of 1985, which was addressed in the PCA using all NDVI images from January and March to October to output principal components. Various tolerance parameters ranging between 10 and 100 were used and segmentation results were examined against a PCA composite and a color-infrared (CIR) Landsat image from August 1985. This method is in contrast to Maxwell (2011), which used Definiens eCognition software to specify different parameters (scale, shape, and compactness) to derive boundaries.

The segmentation process should be further explored to optimally derive segments truly representative of crop field boundaries. For example, Figure 24 shows some segments that appear to cross multiple crop field boundaries (likely same crop type) and also multiple segments are present within the same crop field. It should be noted that as a part of the modified three-tier pesticide exposure method, all crop fields from each dataset (land use survey and Landsat) were dissolved, separately, within each PLSS section - the geographic level of reporting for PUR data. Therefore, although multiple, adjacent segments of the same crop type did not impact pesticide exposure estimation, it is still meaningful to produce resultant segments that truly represent real-world features.

5.1.5 Modified Three-Tier Approach: Pesticide Exposure

A new modified three-tier pesticide exposure methodology was developed, which honored the existing Rull and Ritz (2003) three-tier method through utilizing a tiered approach incorporating land use survey and PLSS section data. However, the modified method introduced tiers derived from Landsat imagery classified according to agricultural crop types, which allowed for determining the independent contribution of Landsat imagery to tiered matching above and beyond land use survey and PLSS data (Tiers 2A and 2C; Figures 27-29).

Implementing the modified three-tier approach demonstrated that most PUR records were claimed by Tier 1 using land use survey crop fields. Results may differ if examining pesticide chemical classes other than organochlorines, organophosphates, and carbamates, time periods beyond 1990, and individual years. Furthermore, as crop types are not perfectly comparable between the PUR and land use survey datasets (e.g. some PUR crop types not found in land use survey), PUR-to-land use survey crosswalk crop assignments may have affected tiered results.

Another methodological consideration is the way in which overlapping crop fields from the land use survey and Landsat data were treated. Rates were calculated by weighting the application rates (pounds of applied pesticides divided by acres of land use survey- or Landsat-derived crop field or PLSS section within each section) by the proportion of the ZCTA or census tract comprised of that particular crop field or section. This weighted average approach takes into account the entire geographic area of each areal unit, irrespective of pesticide treatment. The tiered approach was implemented in such a way that treated Landsat-derived crop fields as boundaries independent of land use survey crop fields without the use of a spatial union to sum applied pounds within overlapping areas. In other words, land use survey and Landsat crop field boundaries matched to PUR data may have intersected each other. This does not affect PLSS section boundaries as the design of the modified three-tier method only matches PUR data to a section when no land use survey or Landsat crop fields are present in a section. It may be appropriate to treat land use survey and Landsat crop fields independently as a particular crop field may exist at one point in time, and be replaced with some other crop field at a later time.

However, these occurrences were infrequent as Landsat modestly contributed tiered matches. Specifically, 3.42% of organochlorine-treated land use survey crop fields were intersected by organochlorine-treated Landsat crop fields (20,022.62 ac; data not shown), followed by 4.82% of carbamate-treated land use survey fields intersecting Landsat fields (32,750.72 ac), and 6.21% of organophosphate-treated land use survey fields intersecting Landsat fields (45,156.36 ac). Future research should explore the optimal way in which to incorporate overlapping crop field boundaries representing multiple time periods.

Another prominent issue was that of sliver polygons, resulting from the intersection of PLSS sections and land use survey or Landsat crop fields. The geographic resolution of PUR data (PLSS section level) necessitated the intersection of these data layers to identify likely locations of treated crop fields when implementing tiered matching. Specifically, if multiple fields of a particular crop type exist within a section, PUR data does not discriminate between which crop field was treated (Goldberg et al. 2007). As crop fields may span multiple sections, sliver polygons were produced as a result of their intersection. One approach to handling sliver polygons is to retain the acreage of the sliver polygon with its source crop field in calculating application rates. Another approach, adopted in this research, was to exclude sliver polygons using shape length and shape area attributes. However, the criteria for exclusion likely excluded smaller crop fields.

The contribution of Landsat was modest across all pesticide chemical classes at Tiers 2A (2.1-2.4%) and 2C (0.1-0.2%), which supports the notion that integrating Landsat remote sensing improved, to a small degree, pesticide exposure assessment through addressing PUR applications that did not match land use survey crop fields. If Landsat was implemented as a tier prior to considering land use survey crop fields, its contribution may have been more pronounced. However, as the Landsat layer was derived from a CSL that used the 1990 Kern County land use survey as a ground truth, it was more appropriate to use the land use survey as Tier 1. In addition, if implemented in a different California county, the contribution of Landsat may have differed due to the prevalence of different crop types. For example, a California county with a higher prevalence of rice crop fields (*N*=1 rice crop field in 1990 Kern County land use survey) may have observed larger tier contributions from Landsat data. Nevertheless, the feasibility component of this study was to modify the existing Rull and Ritz (2003) three-tier approach to evaluate if Landsat could contribute additional crop field location information beyond land use survey data. Integrating Landsat, by way of creating a CSL and classification, was demonstrated to be a feasible analytic addition to the pesticide exposure ascertainment process.

5.1.6 Impact of Areal Aggregation on Annual Pesticide Application Rates

Annual pesticide application rates varied according to areal aggregation and pesticide chemical class. Rates were not significantly different at the ZCTA vs. census tract level, although maximum rates were typically higher at the census tract level. For example, organochlorine-specific annual application rates ranged from 0 to 0.26 lb/ac (median 0.001 lb/ac) at the ZCTA level and from 0 to 1.57 lb/ac (median 0.001 lb/ac) at the census tract level. Differences in application rates according to areal aggregation are a manifestation of the modifiable areal unit problem (MAUP). The delineation of ZCTA and census tract boundaries does not necessarily reflect agricultural crop boundaries, let alone pesticide-treated crop field boundaries and pesticide application practices.

Another important consideration is the geoprocessing of ZCTA boundaries prior to calculating application rates. The calculated ZCTA-level rates reflect pesticide exposure specifically associated with residence within Kern County - by virtue of extracting Kern County PUR records. For example, the full extent of ZCTAs spanning multiple counties, such as 93527, was not considered. ZCTA boundaries were clipped to the Kern County extent for use in weighting application rates. Calculated ZCTA rates represent an ecologic measure of pesticide exposure for individuals residing in both Kern County and a particular ZCTA.

5.1.7 Accuracy Assessment of Rurality

The performance of each rurality metric as a surrogate for pesticide exposure was largely a function of the rurality metric, areal aggregation, pesticide chemical class, and pesticide exposure cutoff. It was hypothesized that RUCA codes, by virtue of incorporating both population and work commuting information, would be both more sensitive and specific in assigning pesticide exposure compared to the GIS gold standard metric. Rurality is an intuitive surrogate measure of pesticide exposure as agricultural pesticide applications occur more frequently in rural areas (Franklin and Worgan 2005).

However, rural ZCTAs and census tracts were typically characterized by lower median annual pesticide application rates compared to their urban counterparts across all chemical classes. These patterns run counter to what was expected, i.e. higher pesticide exposure in rural geographic areas. These results potentially shed light on the distinction between rural and urban areas being unrelated to pesticide application practices in Kern County, which is predominantly rural. It is conceivable that urban areas, as demarcated by RUCA codes and the U.S. Census Bureau metric in Kern County, are actually more rural as compared to urban areas in other counties outside of the Central Valley - by virtue of selecting a predominantly rural study area.

Furthermore, a rurality metric that is not binary and has multiple categories corresponding to different levels of rurality may be more appropriate in trying to capture pesticide exposure. The RHRC presents additional methods to categorize RUCA codes (RHRC 2000). For example, a four-category classification discriminates between areal units that are urban, large rural, small rural, and isolated. In addition, had other pesticide chemical classes been examined, the expected pattern of higher rates in rural geographic areas may have been observed.

Another plausible interpretation stems from the imperfect implementation of the U.S. Census Bureau metric with respect to ZCTAs and census tracts, i.e. handling the large swaths of ZCTAs and census tracts not intersecting Urbanized Areas and Urban

Clusters (Appendix F; Figure F1). Even if a small proportion of a ZCTA or census tract intersected with an Urbanized Area or Urban Cluster, it was classified as urban. Therefore, as fewer ZCTAs and census tracts were classified as rural, there was less opportunity for pesticide-treated crop fields to intersect rural areal units. An area cutoff could have been applied, where an Urbanized Area or Urban Cluster must have intersected a particular proportion of the ZCTA or census tract for it to be classified as urban. However, this approach would have ignored the portion of the areal unit that truly was urban, even if the proportion of the overall areal unit intersecting the Urbanized Area or Urban Cluster was small. The U.S. Census Bureau metric may be more useful when using, for example, individual-level residential locations, rather than areal aggregations. Results regarding the sensitivity and specificity of the U.S. Census Bureau metric should be interpreted with caution.

The extent to which the rurality metrics differed in how Kern County was classified as rural vs. urban was striking. Geographic patterns of rurality were seemingly similar at the ZCTA level - rural ZCTAs using RUCA codes and the U.S. Census Bureau metric were observed in the eastern and western portions of Kern County. However, agreement was poor between the two metrics (*kappa*=0.03). At the census tract level, rurality was more widespread when using RUCA codes - in the eastern, western, and central portions of the county. Rurality, when measured using the U.S. Census Bureau metric, was observed in the northeastern and eastern portions of Kern County. Agreement was also poor between the two metrics (*kappa*=0.04).

In terms of the standard by which to judge satisfactory vs. unsatisfactory sensitivity and specificity, absolute differences from 100% (perfect sensitivity and specificity) and relative differences across areal aggregations, pesticide chemical classes, and pesticide exposure cutoffs were considered. Across both areal aggregations, specificity was superior to sensitivity. This reflects the satisfactory capacity of rurality, whether measured using RUCA codes or the U.S. Census Bureau metric, to correctly identify geographic units truly not exposed to pesticides. At the ZCTA level, RUCA codes were superior to the U.S. Census Bureau metric. The highest specificity for RUCA codes was observed for carbamates (60%) using a cutoff of 0 lb/ac as pesticide-exposed, while the highest specificity for the U.S. Census Bureau metric (51.4%) was observed for all chemical classes using a 75th percentile cutoff. RUCA codes were also superior to the U.S. Census Bureau metric in terms of sensitivity - highest observed for carbamates (42.9%) using a 0 lb/ac cutoff. The highest sensitivity offered by the U.S. Census Bureau metric was observed for organophosphates (35.6%) using a 0 lb/ac cutoff. A larger number of statistically significant differences when comparing the U.S. Census Bureau metric to the GIS gold standard is also indicative of its mediocre performance at the ZCTA level.

At the census tract level, specificity was also consistently higher than sensitivity across all pesticide chemical classes and pesticide exposure cutoffs. Sensitivity and specificity remained relatively constant across all pesticide exposure cutoffs. In contrast to the ZCTA level, the U.S. Census Bureau metric offered superior specificity compared to RUCA codes. The highest specificity using the U.S. Census Bureau metric was observed for organochlorines (97.5%) using a 0 lb/ac cutoff, compared to 77.1% for organochlorines using a $\geq$75th percentile cutoff when with RUCA codes. Across all pesticide chemical classes, sensitivity was poor ($\leq$19.4%). RUCA codes were more sensitive than the U.S. Census Bureau metric- the highest sensitivity was observed for organophosphates (19.4%) using a 0 lb/ac cutoff. The highest sensitivity offered by the U.S. Census Bureau metric was observed for organochlorines (6%) using a 0 lb/ac cutoff.

The prevalence of different pesticide chemical classes used on crops across Kern County impacted the results of the rurality accuracy assessment. RUCA codes were most sensitive and specific for carbamates usage (0 lb/ac cutoff) at the ZCTA level. Interestingly, pesticide-treated crop fields and sections were quite prevalent throughout Kern County, irrespective of chemical class. The largest concentration of treated fields was towards the northwest, reflecting pervasive agricultural practices in this region of the Central Valley. Therefore, although organophosphate usage accounted for the majority of PUR applications and pounds applied in Kern County, carbamate-treated crop fields happened to intersect rural ZCTAs more frequently, and non-treated fields urban ZCTAs, when using the RUCA metric.

On the other hand, when evaluating census tracts, the U.S. Census Bureau metric was most specific for organochlorine usage (0 lb/ac cutoff) and RUCA codes were most sensitive for organophosphate usage (0 lb/ac cutoff). Organochlorines were associated with the fewest median number of pesticide-treated fields intersecting census tracts, both rural and urban, compared to organophosphates and carbamates (Appendix F; Table F2). This was paired with the fact that the U.S. Census Bureau metric classifies the majority of Kern County as urban, which served to increase its specificity with respect to organochlorine usage. The relatively higher prevalence of organophosphates worked to increase the capacity of RUCA codes to accurately classify census tracts as pesticide-exposed/rural, ultimately increasing sensitivity.

The results of the accuracy assessment were also affected by the pesticide chemical classes included the analysis. Different pesticide active ingredients were likely used more frequently for particular crop types, and evaluating only organochlorines, organophosphates, and carbamates may not reflect this pesticide usage in Kern County. For example, excluding other pesticide chemical classes may have misclassified areal units truly pesticide-exposed as not exposed to pesticides (using gold standard), which would have decreased the specificity of both rurality metrics (increased false positive rate). In other words, an areal unit categorized as rural would have been designated as not exposed to pesticides using the current gold standard GIS metric, but would have been designated as pesticide-exposed had other chemical classes been considered.

In a real-world scenario implementing a comparative epidemiologic study using rurality as a surrogate measure of pesticide exposure (assuming low prevalence of pesticide exposure in study population), usage of ZCTA-level RUCA codes (using 0 lb/ac cutoff) would result in less attenuation (i.e. bias towards null hypothesis) in study results due to its superior specificity across all pesticide chemical classes. On the other hand, although usage of census tract-level RUCA codes is associated with relatively high specificity, census tract-level U.S. Bureau Census urban-rural designations were associated with even higher specificity, and would thus result in less attenuation (using 0 lb/ac cutoff). It is important to note that the gain with respect to specificity must be balanced against the impact of sensitivity. Sensitivity was found to be mediocre at the ZCTA level and poor at the census tract level, meaning truly pesticide-exposed areal units were misclassified as not exposed (urban) according to each rurality metric (false negatives). These results are limited in generalizability as the study area was Kern

County, which is predominantly rural; rurality may perform differently as a surrogate measure in different geographic areas across California. In regions with higher pesticide exposure (>10%), sensitivity has a greater impact on study results (Szklo and Nieto 2007).

The methods by which ZCTAs and census tracts are delineated also affected the results of the accuracy assessment. Census tracts, characterized by a larger number of areal units constituting Kern County (N=140) compared to ZCTAs (N=47), may be more homogeneous with respect to urban/rural characteristics. This homogeneity may be more relevant to delineating urban vs. rural areas, reflected in the higher specificity (ability to correctly identify areal units not pesticide-exposed) across all census tract-level rurality metrics when compared to the GIS gold standard. The highest census tract-level specificity was 97.5% compared to 60% at the ZCTA level. This alludes to the notion that census tract aggregations may be delineated in such a way that better determines urban processes lacking agricultural pesticide applications. Furthermore, the overall poor sensitivity (ability to correctly identify areal units exposed to pesticides) observed at the census tract level may also reflect how census tracts are designed in such a way that is not conducive to correctly identifying geographic units truly exposed to pesticides.

Choice of pesticide exposure cutoff (>0 lb/ac, ≥50th percentile, and ≥75th percentile) should be guided by knowledge of pesticide exposures meaningful to the application at hand. Usage of a 0 lb/ac cutoff is the most liberal, expected to result in the highest sensitivity, or capacity to capture all pesticide-exposed areal units. On the other hand, usage of the most conservative cutoff in this study, for example the 75th percentile, would be expected to result in higher specificity, or the capacity to identify all nonexposed areal units. Results may also differ depending on which gold standard pesticide exposure metric is used - for example, a gold standard without Landsat imagery integration. Although PUR-derived, GIS-based pesticide exposure metrics do not directly address all possible routes of pesticide exposure, such as from occupation and diet, they do reflect residential proximity to agricultural pesticides, which has been demonstrated to be significantly associated with within-household pesticide levels (Gunier et al. 2011).

5.2 Feasibility and Informational Gain of Landsat Remote Sensing

The feasibility and utility of integrating Landsat remotely sensed imagery into an existing three-tier pesticide exposure methodology is manifest in the creation of a CSL, classification of additional Landsat imagery, and PUR records matching to Landsat-derived tiers. However, given the constraints of computing power and time, the benefit of incorporating Landsat imagery may be greater at finer spatial resolutions - finer than the ZCTA and census tract levels. By design, annual pesticide application rates were derived by weighting rates according to the proportion of the areal unit comprised of the field or section. Therefore, integrating Landsat imagery may only result in incremental improvements in enhancing pesticide exposure ascertainment when examining ZCTA- and census tract-level rates. Although the weighted average approach to calculating rates represents an average exposure for all individuals residing in a ZCTA or census tract, evaluating such large areal units masks the heterogeneity that exists in pesticide application rates at larger scales. When estimating pesticide exposure at aggregated analysis scales, a PLSS section-only method (Bell et al. 2001; Gunier et al. 2001) or a land use survey and PLSS section method (Rull and Ritz 2003), may produce similar pesticide application rates. To effectively evaluate if Landsat imagery meaningfully improves the spatiotemporal resolution of pesticide exposure estimation, the performance of the modified three-tier method compared to existing methodologies should be assessed at various spatial resolutions (e.g. individual-level residences vs. aggregated analysis scales).

Another methodological approach would have been to classify additional Landsat images from multiple years between 1974 and 1990, not only 1985, which could have improved the informational gain from Landsat data. The modified three-tier method could have incorporated the date of PUR applications in the tiered matching. The availability of Landsat imagery beginning in 1972 allows for this approach. However, earlier Landsat sensors, such as the Multispectral Scanner (MSS), will differ in such characteristics as spatial resolution. This would have to be reconciled against using the 1990 CSL to classify Landsat images, which is constrained to 1990 due to land use survey ground truth data availability. Future research should explore the extent to which classification is affected by using a CSL that differs in spatial resolution to the classified images.

5.3 Alternative Approaches to Integrating Landsat in Pesticide Exposure Estimation

Although the proposed modified three-tier method yielded a modest amount of PUR tier matches using Landsat-derived crop fields, an approach that matches pesticide data to Landsat-derived crop fields (not considering land use surveys) would better address the potential utility of Landsat remote sensing in pesticide exposure estimation. Landsat imagery provides a valuable opportunity to address temporal voids in land use surveys through its moderate temporal resolution (16 to 18 days) in capturing Earth imagery since 1972. Although many agricultural crop types are likely long-standing between years, some agricultural fields are rotated on an annual basis. Crop rotation is implemented to maintain soil fertility via alternation of plant species. For example, a crop rotation schedule may require planting an agricultural parcel of land with a different crop type each year. Sloping land may be subject to excessive soil loss if row crops, such as corn, are grown for many consecutive years. Rotation of corn with sod-based forage crops (e.g. grass) minimizes soil loss (Lerner and Lerner 2008). Therefore, Landsat imagery, when classified accurately, can provide a temporally accurate snapshot of agricultural lands with ≥30 m spatial resolution in delineating crop field boundaries, not otherwise provided with the intermittently updated land use surveys. A formal comparison between the Rull and Ritz (2003) method and a three-tier approach using Landsat-derived crop fields should be implemented to determine the extent to which Landsat imagery can provide a more temporally accurate agricultural landscape at a particular point in time lacking land use surveys – ultimately addressing Landsat's utility in pesticide exposure estimation (i.e. tier 1 PUR matches).

The proposed classification method using a sum of squared differences measure harnesses the temporal variability of NDVI to identify crop types. Typical minimum distance measures classify imagery according to the minimum distance between a pixel value and the mean value of an informational class (Campbell and Wynne 2011). The sum of squared differences measure, as applied in this research, determined the minimum distance (squared) between each segment's median NDVI value and the median NDVI value of each land use class – summed across all months in 1985 with available imagery.

However, other classification methods could be used that can also take into account the temporal variability of NDVI that may also yield accurate results. For example, Wardlow and Egbert (2008) implemented an unsupervised classification approach [Iterative Self-Organizing Data Analysis Technique (ISODATA)] to classify a time series of Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI images of Kansas between March 22 and November 1. Spectral-temporal clusters were generated and assigned to the crop, non-crop, or confused classes via comparing the distribution of the cluster's pixels, cluster means, and visual interpretation of the land cover types using Landsat Enhanced Thematic Mapper Plus (ETM+) imagery. Future research should explore the optimal and most accurate way to classify a temporal series of NDVI images into agricultural crop fields.

5.4 Significance of Results

Demonstrating the feasibility of using remotely sensed data in a GIS-based pesticide exposure metric at cancer data analysis scales to enhance the spatiotemporal resolution of identifying pesticide-applied locations is needed. Quantifying the exact extent of exposure misclassification from using two surrogate measures of rurality compared to a GIS-based pesticide exposure gold standard, as well as determining which measure is superior in terms of accuracy, have never before been performed. The results of this research are specifically relevant to cancer epidemiology - the units of analysis included census tracts and ZIP codes (ZCTAs used to approximate boundaries), which are typical geographic aggregations for cancer registry data used to preserve patient confidentiality (Boscoe et al. 2004; Waller and Gotway 2004). The results shed light on potential exposure misclassification when using rurality-based metrics, and are applicable to ecologic studies utilizing pesticide data aggregated to areal units and individual-level studies using contextual, ecologic metrics.

The results are generalizable to epidemiologic literature examining pesticide exposure in California and other states with similar data. In the absence of data on pesticide applications and land use, understanding which rurality metric most meaningfully captures the processes underlying pesticide exposure - both in terms of the rurality definition and analysis scale - is important to explore. The impact of pesticides on elevating the risk for certain cancers has been established (Alavanja et al. 2004), and research into how to accurately measure pesticide exposure is integral to implementing epidemiologic studies addressing this research topic. Ultimately, this research harnessed GIS tools in order to directly address how the validity of surrogate measures of exposure can directly impact the inferences derived from epidemiologic studies investigating human health outcomes.

5.5 Future Directions

Future research should explore the utility of integrating Landsat into GIS-based pesticide exposure metrics beyond Kern County and at finer spatial scales [e.g. within the 500 m residential buffers implemented by Rull and Ritz (2003)]. A formal comparison between the Rull and Ritz (2003) three-tier method and a Landsat-only pesticide exposure method would be better able to highlight the contribution of Landsat imagery to locating agricultural pesticide applications and ultimately to pesticide exposure estimation. A validity study demonstrating the accuracy of the CSL in discriminating between land use classes when examining imagery from a different geographic area and at a different time point (e.g. creation of an error matrix comparing Landsat-classified crop fields to a ground truth) would highlight the accuracy and generalizability of the CSL. Alternative methods of classification to produce agricultural crop fields using temporal NDVI data should be explored. Investigating the contribution of Landsat in PUR-matching at different points in time would shed light on how Landsat could enhance the temporal resolution of identifying pesticide-treated crop fields. The performance of rurality compared to different GIS-based gold standard pesticide exposure metrics would also be informative. Measuring pesticide exposure as a cumulative measure (e.g. total pounds of applied pesticide) within an areal aggregation as opposed to pesticide exposure density (rate in lb/ac) presented in this research would be valuable to explore as well.

5.6 Summary

The feasibility of incorporating Landsat remotely sensed imagery into a modified GIS-based pesticide exposure metric accommodating cancer data analysis scales was demonstrated. Strengths included the methodological improvement over previous research via objectively harnessing all NDVI spectral information in creating a crop signature library for use in subsequent classification of Landsat imagery, and developing a modified three-tier pesticide exposure method that can be used at other analysis scales. The accuracy of commonly used rurality metrics as pesticide exposure surrogates was assessed, which has never before been researched. RUCA codes offer superior specificity at the ZCTA level while the U.S. Census Bureau urban-rural classification metric offers superior specificity at the census tract level. Accuracy varies according to rurality metric, areal aggregation, pesticide chemical class, and pesticide exposure cutoffs, which should be tailored to specific research applications. Future research should explore the integration of Landsat imagery at finer spatial resolution pesticide exposure methodologies (i.e. individual-level), examine the contribution of a Landsat-only method to estimate pesticide exposure compared to the existing Rull and Ritz (2003) three-tier method, validate the NDVI crop signature library, and evaluate the utility of using different GIS-based pesticide exposure metrics (e.g. cumulative pounds).

REFERENCES

AgroPages. 2013. *Crop Protection Database* 2013 [cited 2013]. Available from http://www.agropages.com/AgroData/.

Alavanja, M. C., J. A. Hoppin, and F. Kamel. 2004. Health effects of chronic pesticide exposure: cancer and neurotoxicity. *Annu Rev Public Health* 25:155-97.

Alavanja, M. C., M. H. Ward, and P. Reynolds. 2007. Carcinogenicity of agricultural pesticides in adults and children. *J Agromedicine* 12 (1):39-56.

Beck, P. S., C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore. 2006. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote sensing of environment* 100 (3):321-334.

Bell, E. M., I. Hertz-Picciotto, and J. J. Beaumont. 2001. A case-control study of pesticides and fetal death due to congenital anomalies. *Epidemiology* 12 (2):148-56.

Blair, A., Zahm, S.H., Cantor, K.P., Stewart, P.A. 1988. Estimating Exposure to Pesticides in Epidemiological Studies of Cancer. In *Biological Monitoring for Pesticide Exposure: Measurement, Estimation, and Risk Reduction*, 38-46: American Chemical Society.

Boscoe, F. P., M. H. Ward, and P. Reynolds. 2004. Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. *Int J Health Geogr* 3 (1):28.

CA.gov. *State of California: Facts* 2013 [cited 2013. Available from http://www.ca.gov/about/facts.html.

Cal-Atlas Geospatial Clearinghouse. 2013 [cited 2013. Available from http://www.atlas.ca.gov/download.html.

Campbell, J. B., and R. H. Wynne. 2011. *Introduction to Remote Sensing*. Fifth ed. New York, NY: The Guilford Press.

CDC (Centers for Disease Control and Prevention). *Fourth National Report on Human Exposure to Environmental Chemicals* 2009 [cited 2013. Available from http://www.cdc.gov/exposurereport/pdf/FourthReport.pdf.

———. *Cancer* 2012 [cited 2013. Available from http://www.cdc.gov/chronicdisease/resources/publications/AAG/dcpc.htm.

CDOC (California Department of Conservation). *FMMP - Mapping Procedures* 2013 [cited 2013. Available from http://www.conservation.ca.gov/dlrp/fmmp/mccu/Pages/making_map.aspx.

CDPR (California Department of Pesticide Regulation). *Appendix C: California's Pesticide Use Report An Assessment of Spatial Data Quality* 2000a [cited 2013. Available from http://www.cdpr.ca.gov/docs/pur/appendix_c_dataq_ldr.pdf.

———. *DPR Pesticide Use Reporting: An Overview of California's Unique Full Reporting System* 2000b [cited 2013. Available from http://www.cdpr.ca.gov/docs/pur/purovrvw/ovr52000.pdf.

———. 2002. Pesticide Use Report Data User Guide & Documentation.

———. *Pesticide Use Reporting* 2011a [cited 2013. Available from http://www.cdpr.ca.gov/docs/pur/pur11rep/lbsby_co_11.pdf.

———. *Top 5 Sites* 2011b [cited 2013. Available from http://www.cdpr.ca.gov/docs/pur/pur11rep/top_5_sites_ais_lbs11.pdf.

———. *Pesticide Use Reporting (PUR)* 2013 [cited 2013. Available from http://www.cdpr.ca.gov/docs/pur/purmain.htm.

CDWR (California Department of Water Resources). *Land Use Survey Overview* 2013 [cited 2013. Available from http://www.water.ca.gov/landwateruse/lusrvymain.cfm.

Chander, G., B. L. Markham, and D. L. Helder. 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote sensing of environment* 113 (5):893-903.

Chavez, P. S. 1996. Image-based atmospheric corrections-revisited and improved. *Photogrammetric engineering and remote sensing* 62 (9):1025-1035.

Clark Labs. 2013. IDRISI Selva. Worcester, Massachusetts, United States.

Cockburn, M., P. Mills, X. Zhang, J. Zadnick, D. Goldberg, and B. Ritz. 2011. Prostate cancer and ambient pesticide exposure in agriculturally intensive areas in California. *Am J Epidemiol* 173 (11):1280-8.

Cordier, S., T. B. Le, P. Verger, D. Bard, C. D. Le, B. Larouze, M. C. Dazza, T. Q. Hoang, and L. Abenhaim. 1993. Viral infections and chemical exposures as risk factors for hepatocellular carcinoma in Vietnam. *Int J Cancer* 55 (2):196-201.

Costello, S., M. Cockburn, J. Bronstein, X. Zhang, and B. Ritz. 2009. Parkinson's disease and residential exposure to maneb and paraquat from agricultural applications in the central valley of California. *Am J Epidemiol* 169 (8):919-26.

Dich, J., S. H. Zahm, A. Hanberg, and H. O. Adami. 1997. Pesticides and cancer. *Cancer Causes Control* 8 (3):420-43.

Economic Research Service. *Overview* 2012 [cited 2013. Available from http://www.ers.usda.gov/data-products/agricultural-productivity-in-the-us.aspx.

EPA (Environmental Protection Agency). *Pesticide issues in the works: pesticide volatilization* 2009 [cited 2013. Available from http://www.epa.gov/pesticides/about/intheworks/volatilization.htm.

———. *Pesticides Industry Sales and Usage: 2006 and 2007 Market Estimates* 2011 [cited 2013. Available from http://www.epa.gov/opp00001/pestsales/07pestsales/market_estimates2007.pdf.

———. *Pesticides* 2012 [cited 2013. Available from http://www.epa.gov/pesticides/.

Esri. 2013. ArcGIS 10.1. Redlands, California, United States.

Ezzat, S., M. Abdel-Hamid, S. A. Eissa, N. Mokhtar, N. A. Labib, L. El-Ghorory, N. N. Mikhail, A. Abdel-Hamid, T. Hifnawy, G. T. Strickland, and C. A. Loffredo. 2005. Associations of pesticides, HCV, HBV, and hepatocellular carcinoma in Egypt. *Int J Hyg Environ Health* 208 (5):329-39.

Franklin, C., and J. Worgan. 2005. *Occupational and Residential Exposure Assessment for Pesticides*. Hoboken, NJ: Wiley.

Gatto, N. M., M. Cockburn, J. Bronstein, A. D. Manthripragada, and B. Ritz. 2009. Well-water consumption and Parkinson's disease in rural California. *Environ Health Perspect* 117 (12):1912-8.

Goldberg, D. W., X. Zhang, J. C. Marusek, J. P. Wilson, B. Ritz, and M. G. Cockburn. 2007. Development of an automated pesticide exposure analyst for California's central valley. Paper read at Proc Urban Regional Info Syst Assoc GIS Public Health Conf, New Orleans.

Greene, S. A., and R. P. Pohanish. 2005. *Sittig's handbook of pesticides and agricultural chemicals*. Norwich, N.Y.: William Andrew Pub.

Grubesic, T. H., and T. C. Matisziw. 2006. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *Int J Health Geogr* 5:58.

Gunier, R. B., M. E. Harnly, P. Reynolds, A. Hertz, and J. Von Behren. 2001. Agricultural pesticide use in California: pesticide prioritization, use densities, and population distributions for a childhood cancer study. *Environ Health Perspect* 109 (10):1071-8.

Gunier, R. B., M. H. Ward, M. Airola, E. M. Bell, J. Colt, M. Nishioka, P. A. Buffler, P. Reynolds, R. P. Rull, A. Hertz, C. Metayer, and J. R. Nuckols. 2011. Determinants of agricultural pesticide concentrations in carpet dust. *Environ Health Perspect* 119 (7):970-6.

Jacquez, G. M. 2004. Current practices in the spatial analysis of cancer: flies in the ointment. *Int J Health Geogr* 3 (1):22.

Lee, P. C., Y. Bordelon, J. Bronstein, and B. Ritz. 2012. Traumatic brain injury, paraquat exposure, and their relationship to Parkinson disease. *Neurology* 79 (20):2061-6.

Lee, P. C., S. L. Rhodes, J. S. Sinsheimer, J. Bronstein, and B. Ritz. 2013. Functional paraoxonase 1 variants modify the risk of Parkinson's disease due to organophosphate exposure. *Environ Int* 56:42-7.

Lerner, K. L., and B. W. Lerner. 2008. Crop Rotation. In *The Gale Encyclopedia of Science*, 1187-1190. Detroit.

Lippitt, C. D., L. L. Coulter, M. Freeman, J. Lamantia-Bishop, W. Pang, and D. A. Stow. 2012. The effect of input data transformations on object-based image analysis. *Remote Sensing Letters* 3 (1):21-29.

Manthripragada, A. D., S. Costello, M. G. Cockburn, J. M. Bronstein, and B. Ritz. 2010. Paraoxonase 1, agricultural organophosphate exposure, and Parkinson disease. *Epidemiology* 21 (1):87-94.

Marusek, J. C., M. G. Cockburn, P. K. Mills, and B. R. Ritz. 2006. Control selection and pesticide exposure assessment via GIS in prostate cancer studies. *Am J Prev Med* 30 (2 Suppl):S109-16.

Mather, P. M., and M. Koch. 2011. Chapter 7: Filtering Techniques. In *Computer Processing of Remotely-Sensed Images: An Introduction, 4th Edition*
Oxford, UK: Wiley Blackwell.

Maxwell, S., J. Meliker, and P. Goovaerts. 2010a. Use of land surface remotely sensed satellite and airborne data for environmental exposure assessment in cancer research. *J Expo Sci Environ Epidemiol* 20 (2):176-85.

Maxwell, S. K. 2010. Generating land cover boundaries from remotely sensed data using object-based image analysis: overview and epidemiological application. *Spat Spatiotemporal Epidemiol* 1 (4):231-7.

———. 2011. Downscaling Pesticide Use Data to the Crop Field Level in California Using Landsat Satellite Imagery: Paraquat Case Study. *Remote Sensing* 3 (9):1805-1816.

Maxwell, S. K., M. Airola, and J. R. Nuckols. 2010b. Using Landsat satellite data to support pesticide exposure assessment in California. *Int J Health Geogr* 9:46.

Ministry of Agriculture. 2013. *Pesticide Application Equipment*. British Columbia 2013 [cited April 3 2013]. Available from http://www.agf.gov.bc.ca/pesticides/f_2.htm#3.

Montello, D. R. 2001. Scale in Geography. In *International Encyclopedia of the Social & Behavioral Sciences*, ed. N. J. Smelser, Baltes, P.B., 13501-13504: Pergamon Press.

National Atlas. *The Public Land Survey System (PLSS)* 2013 [cited 2013. Available from http://www.nationalatlas.gov/articles/boundaries/a_plss.html.

NCI (National Cancer Institute). *SEER Stat Fact Sheets: All Sites* 2012 [cited 2013. Available from http://seer.cancer.gov/statfacts/html/all.html#incidence-mortality.

NPR (National Public Radio). *California's Central Valley* 2002 [cited 2013. Available from http://www.npr.org/programs/atc/features/2002/nov/central_valley/.

Nuckols, J. R., R. B. Gunier, P. Riggs, R. Miller, P. Reynolds, and M. H. Ward. 2007. Linkage of the California Pesticide Use Reporting Database with spatial land use data for exposure assessment. *Environ Health Perspect* 115 (5):684-9.

Nuckols, J. R., M. H. Ward, and L. Jarup. 2004. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environ Health Perspect* 112 (9):1007-15.

O'Sullivan, D. O., and D. J. Unwin. 2010. *Geographic Information Analysis*. 2nd ed: John Wiley & Sons.

Oates, L., and M. Cohen. 2011. Assessing diet as a modifiable risk factor for pesticide exposure. *Int J Environ Res Public Health* 8 (6):1792-804.

Persson, E. C., B. I. Graubard, A. A. Evans, W. T. London, J. P. Weber, A. Leblanc, G. Chen, W. Lin, and K. A. McGlynn. 2012. Dichlorodiphenyltrichloroethane and risk of hepatocellular carcinoma. *Int J Cancer*.

Pickle, L. W., L. A. Waller, and A. B. Lawson. 2005. Current practices in cancer spatial data analysis: a call for guidance. *Int J Health Geogr* 4 (1):3.

Reynolds, P., J. Von Behren, R. B. Gunier, D. E. Goldberg, M. Harnly, and A. Hertz. 2005. Agricultural pesticide use and childhood cancer in California. *Epidemiology* 16 (1):93-100.

Reynolds, P., J. Von Behren, R. B. Gunier, D. E. Goldberg, A. Hertz, and M. E. Harnly. 2002. Childhood cancer and agricultural pesticide use: an ecologic study in California. *Environ Health Perspect* 110 (3):319-24.

RHRC (Rural Health Research Center). *Rural-Urban Commuting Area Codes (RUCAs)* 2000 [cited. Available from http://depts.washington.edu/uwruca/.

Ritz, B., and S. Costello. 2006. Geographic model and biomarker-derived measures of pesticide exposure and Parkinson's disease. *Ann N Y Acad Sci* 1076:378-87.

Ritz, B., and R. P. Rull. 2008. Assessment of environmental exposures from agricultural pesticides in childhood leukaemia studies: challenges and opportunities. *Radiat Prot Dosimetry* 132 (2):148-55.

Ritz, B. R., A. D. Manthripragada, S. Costello, S. J. Lincoln, M. J. Farrer, M. Cockburn, and J. Bronstein. 2009. Dopamine transporter genetic variants and pesticides in Parkinson's disease. *Environ Health Perspect* 117 (6):964-9.

Roberts, E. M., P. B. English, J. K. Grether, G. C. Windham, L. Somberg, and C. Wolff. 2007. Maternal residence near agricultural pesticide applications and autism spectrum disorders among children in the California Central Valley. *Environ Health Perspect* 115 (10):1482-9.

Rothman, K. J., S. Greenland, and T. L. Lash. 2008. *Modern epidemiology*: Lippincott Williams & Wilkins.

Rull, R. P., R. Gunier, J. Von Behren, A. Hertz, V. Crouse, P. A. Buffler, and P. Reynolds. 2009. Residential proximity to agricultural pesticide applications and childhood acute lymphoblastic leukemia. *Environ Res* 109 (7):891-9.

Rull, R. P., and B. Ritz. 2003. Historical pesticide exposure in California using pesticide use reports and land-use surveys: an assessment of misclassification error and bias. *Environ Health Perspect* 111 (13):1582-9.

Rull, R. P., B. Ritz, and G. M. Shaw. 2006a. Neural tube defects and maternal residential proximity to agricultural pesticide applications. *Am J Epidemiol* 163 (8):743-753.

Rull, R. P., B. Ritz, and G. M. Shaw. 2006b. Validation of self-reported proximity to agricultural crops in a case-control study of neural tube defects. *J Expo Sci Environ Epidemiol* 16 (2):147-55.

Rural Assistance Center. *What is Rural? Frequently Asked Questions* 2012 [cited 2013. Available from http://www.raconline.org/topics/ruraldef/ruraldeffaq.php#principal.

SAS. 2013. SAS. Cary, North Carolina, United States.

Strahler, A. H., C. E. Woodcock, and J. A. Smith. 1986. On the nature of models in remote sensing. *Remote sensing of environment* 20 (2):121-139.

Szklo, M., and F. Nieto. 2007. Epidemiology: beyond the basics. Johns and Bartlett Publishers. *Inc. Maryland* 2.

U.S. Census Bureau. *Urban and Rural Classification* 2000 [cited 2013. Available from http://www.census.gov/geo/reference/urban-rural.html.

———. *TIGER Products* 2013 [cited 2013. Available from http://www.census.gov/geo/maps-data/data/tiger.html.

USDA (U.S. Department of Agriculture). *Farms, Land in Farms, Value of Land and Buildings, and Land Use: 2007 and 2002* 2007a [cited 2013. Available from http://www.agcensus.usda.gov/Publications/2007/Full_Report/Volume_1,_Chapter_2_County_Level/California/st06_2_008_008.pdf.

———. *Fertilizers and Chemicals Applied: 2007 and 2002* 2007b [cited 2013. Available from http://www.agcensus.usda.gov/Publications/2007/Full_Report/Volume_1,_Chapter_2_County_Level/California/st06_2_042_042.pdf.

———. *Selected Crops Harvested: 2007* 2007c [cited 2013. Available from http://www.agcensus.usda.gov/Publications/2007/Full_Report/Volume_1,_Chapter_2_County_Level/California/st06_2_025_025.pdf.

———. *Rural-Urban Commuting Area Codes: Overview* 2012 [cited 2013. Available from http://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx#.UWSs-kprYRw.

———. *Four Band Digital Imagery: Information Sheet* 2013a [cited 2013. Available from www.fsa.usda.gov/Internet/FSA_File/fourband_info_sheet_2013.pdf.

———. *State Fact Sheets* 2013b [cited 2013. Available from http://www.ers.usda.gov/data-products/state-fact-sheets/state-data.aspx?StateFIPS=06&StateName=California#.UWSfVkprYRw.

USGS (U.S. Geological Survey). *NDVI, the Foundation for Remote Sensing Phenology* 2011 [cited 2013. Available from http://phenology.cr.usgs.gov/ndvi_foundation.php.

———. *Landsat Processing Details* 2013a [cited 2013. Available from http://landsat.usgs.gov/Landsat_Processing_Details.php

———. *Landsat: A Global Land-Imaging Mission* 2013b [cited 2013. Available from http://landsat.usgs.gov/.

———. *Path/Row Shapefiles* 2013c [cited 2013. Available from http://landsat.usgs.gov/tools_wrs-2_shapefile.php.

Vassiliou, A., M. Boulianne, and J. Blais. 1988. On the application of averaging median filters in remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on* 26 (6):832-838.

Waller, L. A., and C. A. Gotway. 2004. *Applied spatial statistics for public health data*: Wiley-Interscience.

Wang, A., S. Costello, M. Cockburn, X. Zhang, J. Bronstein, and B. Ritz. 2011. Parkinson's disease risk from ambient exposure to pesticides. *Eur J Epidemiol* 26 (7):547-55.

Ward, M. H., J. Lubin, J. Giglierano, J. S. Colt, C. Wolter, N. Bekiroglu, D. Camann, P. Hartge, and J. R. Nuckols. 2006. Proximity to crops and residential exposure to agricultural herbicides in iowa. *Environ Health Perspect* 114 (6):893-7.

Ward, M. H., J. R. Nuckols, S. J. Weigel, S. K. Maxwell, K. P. Cantor, and R. S. Miller. 2000. Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System. *Environ Health Perspect* 108 (1):5-12.

Wardlow, B. D., and S. L. Egbert. 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US Central Great Plains. *Remote sensing of environment* 112 (3):1096-1116.

Wood, A. 2013. *Compendium of Pesticide Common Names* 2010 [cited 2013]. Available from http://www.alanwood.net/pesticides/class_pesticides.html

YCEO (Yale Center for Earth Observatoin). *Yale Guide to Landsat 8 Image Processing* 2013 [cited 2013. Available from http://www.yale.edu/ceo/Documentation/Landsat%208%20image%20processing.pdf

APPENDIX A: PESTICIDE DATABASE

Table A1 Organochlorine pesticides[1]

Pesticide name	CDPR chemical code
Aldrin	9
Allidochlor	114
Chlordane	130
Chlorobenzilate	132
Chloroneb	135
DCPA	179
Dalapon	180
TDE	184
DDT	186
Dichlone	202
Dieldrin	210
Endosulfan	259
Endrin	262
Heptachlor	317
Dicofol	346
Chlordecone	347
Gamma-HCH	359
Methoxychlor	384
Mirex	402
Quintozene	464
Pentachlorophenol	465
Dienochlor	468
Ethyl-DDD	472
Tetradifon	581
Toxaphene	594
Trichlorobenzoic Acid	602
Azacosterol dihydrochloride	2026
Acetochlor	2349
HCH	5835

[1] Data from Dich et al. (1997); Gunier et al. (2001); Alavanja, Hoppin, and Kamel (2004); Greene and Pohanish (2005); Rull et al. (2006a, 2009); Wood (2010); and AgroPages (2013)

Table A2 Organophosphate pesticides[1]

Pesticide name	CDPR chemical code
Temephos	1
Monocrotophos	52

Table A2 continued

Pesticide name	CDPR chemical code
Fenthion	63
Bensulide	70
Dicrotophos	72
Trichlorfon	88
Carbophenothion	110
Crotoxyphos	140
Coumaphos	165
Fensulfothion	181
Dichlorvos	187
Tribufos	190
Dioxathion	192
Diazinon	198
Dimethoate	216
Disulfoton	230
Chlorpyrifos	253
Fonofos	254
Butonate	255
EPN	263
Ethion	268
Famphur	282
Tetrachlorvinphos	305
2,4-DEP	306
Azinphos methyl	314
Phosmet	335
Malathion	367
Oxydemeton-methyl	382
Methyl parathion	394
Ethoprophos	404
Naled	418
Schradan	446
Parathion	459
Phorate	478
Phosalone	479
Mevinphos	480
Phosphamidon	482
Fenchlorphos	517
Crufomate	519
Sulfotep	558
Demeton	566
TEPP	577
Dichlofenthion	614

Table A2 continued

Pesticide name	CDPR chemical code
Phosacetim	1523
Ethephon	1626
Leptophos	1676
Acephate	1685
Methidathion	1689
Methamidophos	1697
Dialifos	1799
Glyphosate-isopropylammonium	1855
Fospirate	1856
Fenamiphos	1857
Fosamine ammonium	1921
Edifenphos	1964
Sulprofos	2006
Profenofos	2042
Propetamphos	2122
Isofenphos	2194
Fosetyl-al	2210
Pirimiphos-methyl	2217
Glyphosate-sesquisodium	2275
Isazofos	2282
Omethoate	2285
Glyphosate-trimesium	2327
Isocarbophos	2414
Butathiofos	2433
Chlorpyrifos-methyl	2468
Chlorthiophos	2469
Fenitrothion	2520
Pirimiphos-ethyl	2781
Terbufos	2925
Thionazin	2939
Glyphosate	2997
Triazophos	3543
Vamidothion	3544
Glufosinate-ammonium	3946
Azinphos-ethyl	4053
Demeton-methyl	4063
Paraoxon	4082
Prothiofos	4094
Trichloronate	5001
Chlorethoxyfos	5106

Table A2 continued

Pesticide name	CDPR chemical code
Tebupirimfos	5122
Glyphosate-diammonium	5810

[1] Data from Dich et al. (1997); Gunier et al. (2001); Alavanja, Hoppin, and Kamel (2004); Greene and Pohanish (2005); Rull et al. (2006a, 2009); Wood (2010); and AgroPages (2013)

Table A3 Carbamate pesticides[1]

Pesticide name	CDPR chemical code
Terbucarb	51
Barban	55
Propoxur	62
Bufencarb	91
Carbaryl	105
Carbofuran	106
Formetanate hydrochloride	111
Bcpc	141
Propham	339
Methiocarb	375
Methomyl	383
Aldicarb	575
Pebulate	590
Mexacarbate	623
Phenmedipham	675
Dichlormate	690
Karbutilate	691
Benomyl	1552
Thiophanate	1684
Thiophanate-methyl	1696
Desmedipham	1748
Pirimicarb	1875
Oxamyl	1910
Bendiocarb	1924
Propamocarb	2147
Carbendazim	2176
Carbosulfan	2182
Butoxycarboxim	2201
Thiodicarb	2202
Aldoxycarb	2265

Table A3 continued

Pesticide name	**CDPR chemical code**
Fenoxycarb	2283
Cimectacarb	2345
Aminocarb	2435
Thiofanox	2938
Trimethacarb	2962
Ammonium carbamate	3041
Propamocarb hydrochloride	4022
Dioxacarb	4067
Promecarb	4092
Swep	4098
Asulam	5076
Pyraclostrobin	5759
Iprovalicarb	5938

[1] Data from Dich et al. (1997); Gunier et al. (2001); Alavanja, Hoppin, and Kamel (2004); Greene and Pohanish (2005); Rull et al. (2006a, 2009); Wood (2010); and AgroPages (2013)

APPENDIX B: PESTICIDE USE REPORT PROCESSING

Table B1 Pesticide Use Report (PUR) logic checks[1,2]

Logic check	Definition (1974-1989)	Definition (1990)	Action taken
1. Duplicates	Using county code, acres treated, product number, AI code, pounds of applied AI, application date, and commodity code	Using county code, use number, grower ID, site location ID, acres planted, acres treated, product number, AI code, pounds of applied AI, application date, and commodity code	Kept first record
2. Spatially Inconsistent County	CO-MTRS outside of county boundary	CO-MTRS outside of county boundary	Excluded
3. Inconsistent County Code	N/A	First two digits of grower ID does not match county code	Excluded
4. Missing agricultural field location IDs	N/A	Missing grower ID, site location ID, or CO-MTRS	Excluded
5. Inconsistent CO-MTRS for location	N/A	Given a grower ID and site location ID, different CO-MTRS for different PUR records	Excluded
6. Inconsistent acres planted	N/A	Given a grower ID, site location ID, CO-MTRS, and commodity code, different planted acres	Excluded
7. Treated acres greater than planted acres	N/A	Treated acres greater than planted acres	Excluded

1 Data adapted from CDPR (2000a)

2 Logic checks 3 through 7 were not applied to PUR data extracted between 1974 and 1989 due to missing variables. These logic checks were adapted from CDPR (2000a), which were designed to check PUR data beginning in 1990.

Table B2 Pesticide Use Reports between 1974 and 1989: logic checks[1,2]

Year	PUR records (*N*)	Agricultural use [*N* (%)]	Logic check [*N* (%)]	
			1	2[3]
1974	552,244	433,291 (78.5%)	416,238 (96.1%)	398,377 (91.9%)
1975	583,457	447,837 (76.8%)	394,528 (88.1%)	379,419 (84.7%)
1976	569,142	434,885 (76.4%)	416,626 (95.8%)	397,220 (91.3%)
1977	611,351	472,164 (77.2%)	431,371 (91.4%)	409,957 (86.8%)
1978	476,981	363,844 (76.3%)	347,916 (95.6%)	326,429 (89.7%)
1979	689,568	531,559 (77.1%)	462,887 (87.1%)	436,380 (82.1%)
1980	619,809	454,306 (73.3%)	436,565 (96.1%)	413,197 (91.0%)
1981	691,734	503,078 (72.7%)	481,823 (95.8%)	453,613 (90.2%)
1982	662,702	465,068 (70.2%)	444,310 (95.5%)	420,345 (90.5%)
1983	724,774	464,274 (64.1%)	445,556 (96.0%)	422,128 (90.9%)
1984	832,385	542,628 (65.2%)	521,387 (96.1%)	498,076 (91.8%)
1985	929,918	522,256 (56.2%)	501,877 (96.1%)	482,042 (92.3%)
1986	1,021,166	554,470 (54.3%)	531,502 (95.9%)	504,596 (91.0%)
1987	1,072,329	591,611 (55.2%)	566,131 (95.7%)	535,457 (90.5%)
1988	1,092,688	615,188 (56.3%)	587,519 (95.5%)	552,262 (89.8%)
1989	1,305,573	599,535 (45.9%)	568,657 (94.8%)	539,582 (90.0%)

1 Data from CDPR (2013)
2 Logic checks were sequentially applied to data. Row percentages for logic check 2 use the number of agricultural PURs as the denominator.
3 By design, PURs with an invalid CO-MTRS are excluded at logic check 2.

Table B3 Pesticide Use Reports in 1990: logic checks[1]

PUR records [*N*]	Agricultural use [*N* (%)]	Logic check [N (%)][2]			
2,657,840	2,157,190 (81.2%)	1	2	3	4
		2,092,940 (97.0%)	1,937,646 (89.8%)	1,860,144 (86.2%)	1,859,988 (86.2%)
		5	6	7	
		1,751,727 (81.2%)	1,574,012 (73.0%)	1,540,315 (71.4%)	

1 Data from CDPR (2013)
2 Logic checks were sequentially applied to data. Row percentages for the logic checks use the number of agricultural PURs as the denominator.

Table B4: PUR Outliers[1]

Year	Agricultural use (*N*)	Chemical class [*N* (%)]	Outlier 1[2] [*N* (%)]	Outlier 2[3] [*N* (%)]	Outlier 3[4] [*N* (%)]
1974	398,377	194,488 (48.8%)	3 (0.002%)	126 (0.1%)	…
1975	379,419	182,453 (48.1%)	8 (0.004%)	172 (0.1%)	…
1976	397,220	182,132 (45.9%)	14 (0.008%)	164 (0.1%)	…
1977	409,957	194,199 (47.4%)	17 (0.01%)	189 (0.1%)	…
1978	326,298	149,823 (45.9%)	3 (0.002%)	89 (0.1%)	…
1979	436,378	188,438 (43.2%)	61 (0.04%)	150 (0.1%)	…
1980	413,195	174,153 (42.1%)	326 (0.19%)	419 (0.2%)	…
1981	453,613	193,440 (42.6%)	37 (0.01%)	168 (0.1%)	…
1982	420,344	166,253 (39.6%)	26 (0.02%)	119 (0.1%)	…
1983	422,127	163,010 (38.6%)	22 (0.02%)	108 (0.1%)	…
1984	498,075	198,488 (39.9%)	17 (0.01%)	199 (0.1%)	…
1985	482,041	189,742 (39.4%)	25 (0.01%)	219 (0.1%)	…
1986	504,594	198,915 (39.4%)	43 (0.02%)	265 (0.1%)	…
1987	535,456	215,810 (40.3%)	21 (0.01%)	218 (0.1%)	…
1988	552,261	234,047 (42.4%)	36 (0.02%)	158 (0.7%)	…
1989	539,582	226,483 (42.0%)	35 (0.02%)	168 (0.1%)	…
1990	1,540,315	357,930 (23.2%)	89 (0.02%)	699 (0.2%)	1,765 (0.5%)

1 Data from CDPR (2013)
2 Outlier 1 refers to application rates >200 lb/ac (>1,000 lb/ac if fumigation).
3 Outlier 2 refers to application rates >50 times the median rate for all uses of that pesticide product, commodity code, unit type, and record type.
4 Outlier 3 refers to identification via a neural network.

APPENDIX C: LANDSAT MOSAICS, 1990

Landsat band 3 (red) and band 4 (near infrared) images for Paths 41-42 and Rows 35-36 were mosaicked for the months between January and October 1990. The following two mosaics are examples of the radiometrically corrected (to at-sensor reflectance) images contributing to the crop signature library.

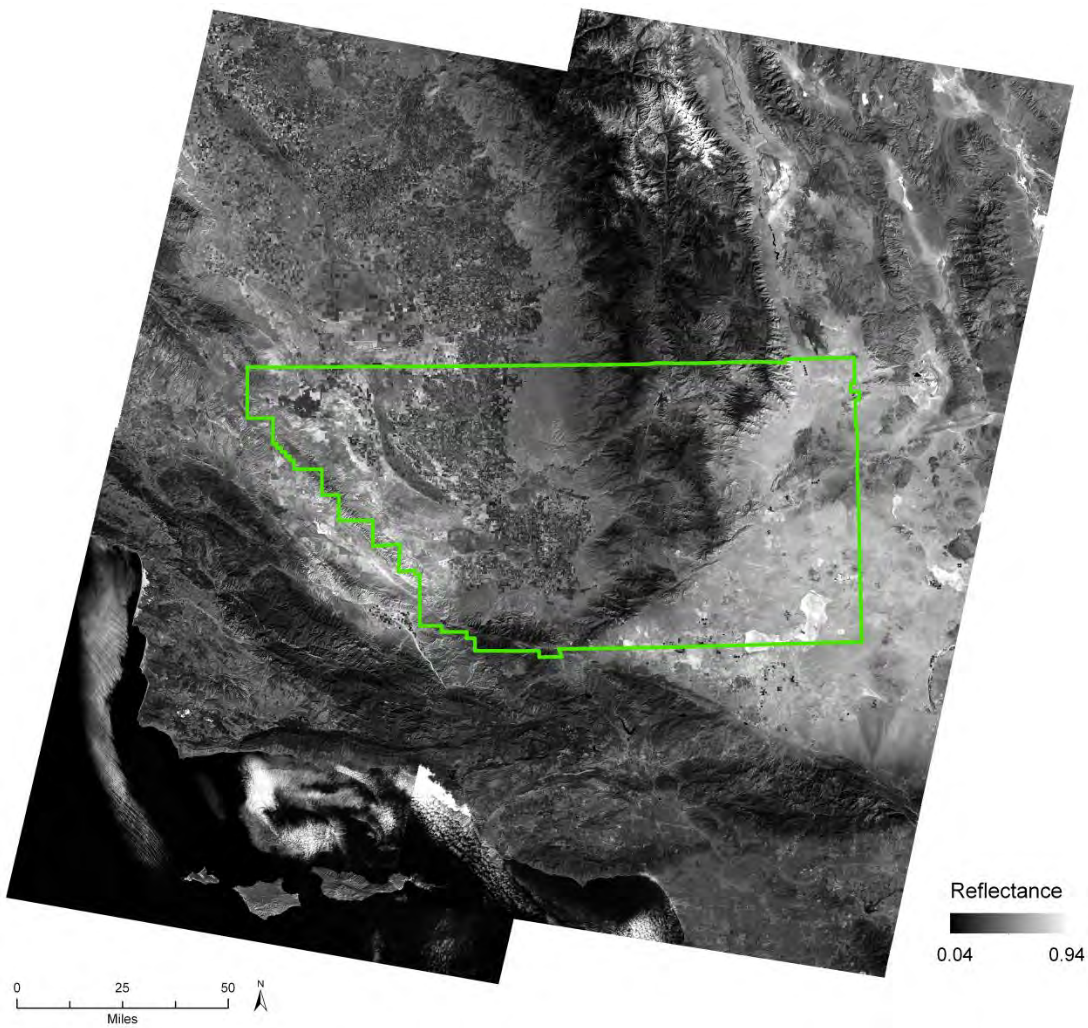


Figure C1 Landsat mosaic (band 3), Paths 41-42 and Rows 35-36, from October 1990 (Data from U.S. Census Bureau 2013; and USGS 2013b)

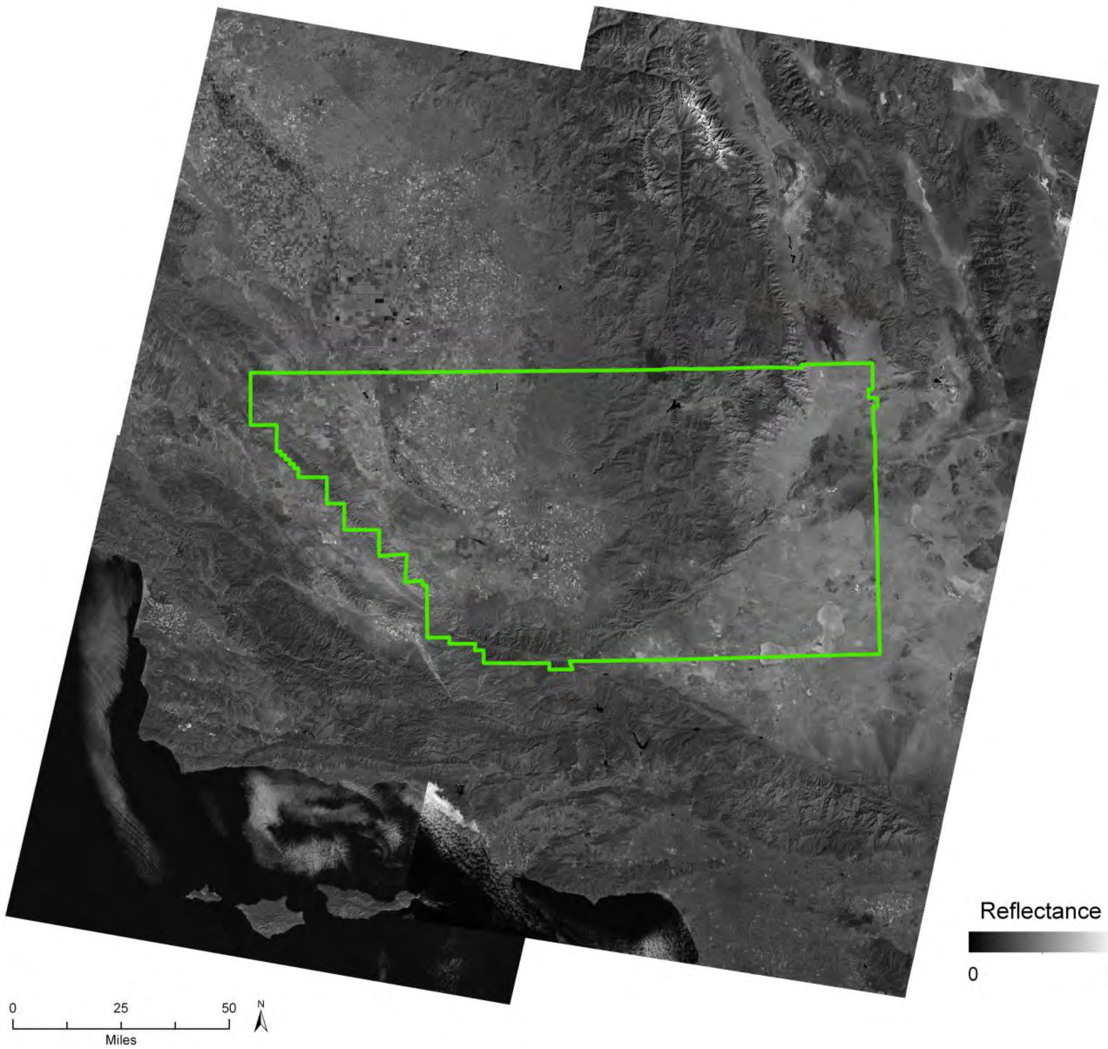


Figure C2 Landsat mosaic (band 4), Paths 41-42 and Rows 35-36, from October 1990 (Data from U.S. Census Bureau 2013; and USGS 2013b)

APPENDIX D: CROP SIGNATURE LIBRARY

Boxplots are presented for agricultural land use classes included in the crop signature library (CSL) via stratified random sampling (SRS) (*N*=55). Agricultural land use classes with few SRS samples will not show distinct boxplot features (e.g. avocado). Broad land use classes, such as without subclass designations in the 1990 Kern County land use survey (e.g. class=F, field crop) are included. Refer to Table D1 for land use class strata sample sizes.

Table D1 Sampled land use class polygons in crop signature library[1,2]

Land use class	**SRS samples (*n*)**	**Land use class**	**SRS samples (*n*)**
Alfalfa	30	Vacant	18
Almond	30	Olive	17
Apple	30	Storage	17
Bean (dry)	30	Pepper	16
Carrot	30	Sweet potato	16
Corn	30	Apricot	14
Cotton	30	Miscellaneous truck	14
Farmstead	30	Idle-new lands prepared for crops	12
Feed lot	30	Unpaved area	12
Field crop	30	Grain sorghum	11
Flowers and nursery	30	Industrial	10
Freeway	30	Fruit and vegetable cannery	9
Grain and hay crop	30	Airport runway	7
Idle-cropped in past year	30	Asparagus	7
Lawn area: irrigated	30	Cemetery: irrigated	7
Lettuce	30	Fig	7
Melon, squash, cucumber	30	Turf farm	7
Mixed pasture	30	Grapefruit	6
Native vegetation	30	Jojoba	6
Onion and garlic	30	Cole crop	4
Orange	30	Prune	4

Table D1 continued

Land use class	SRS samples (*n*)	Land use class	SRS samples (*n*)
Peach and nectarine	30	Cabbage	2
Pistachio	30	Cherry	2
Plum	30	Municipal auditorium	2
Potato	30	Native pasture	2
Residence: 3-4 houses/ac	30	Oil refinery	2
Safflower	30	School	2
Sugar beet	30	Avocado	1
Tomato	30	Bushberry	1
Urban	30	Commercial	1
Vineyard	30	Idle	1
Water surface	30	Manufacturing	1
Lemon	29	Miscellaneous establishment	1
Kiwi	27	Miscellaneous subtropical fruit	1
Walnut	26	Motel	1
Miscellaneous deciduous	24	Paved area	1
Truck, nursery, and berry crop	24	Pea	1
Bean (green)	23	Pear	1
Extractive industry	22	Rice	1
Dairy	21	Urban: residential	1
Sudan	21		

1 Data from CDWR (2013)

2 A total of 1,423 samples across all 81 land use classes included in the crop signature library were included.

Table D2 Monthly NDVI values by land use strata[1,2]

Land use class	NDVI values (*N*)	Land use class	NDVI values (*N*)
Native vegetation	149,648	Grain sorghum	4,728
Pistachio	26,878	Dairy	4,118
Urban	26,540	Sweet potato	3,846
Safflower	20,960	Fig	3,757
Almond	19,746	Vacant	3,405
Cotton	17,531	Jojoba	3,271

Table D2 continued

Land use class	**NDVI values (*N*)**	**Land use class**	**NDVI values (*N*)**
Grain and hay crop	16,031	Asparagus	3,232
Apple	14,798	Kiwi	3,050
Idle-cropped in past year	14,316	Miscellaneous truck	3,022
Vineyard	13,556	Apricot	2,992
Lettuce	13,154	Airport runway	2,464
Olive	12,744	Rice	2,427
Sugar beet	11,620	Farmstead	2,199
Potato	11,191	Storage	1,816
Tomato	11,070	Grapefruit	1,711
Lemon	10,706	Industrial	1,648
Freeway	10,702	Unpaved area	1,519
Alfalfa	10,394	Municipal auditorium	1,232
Bean (dry)	10,359	Turf farm	1,153
Water surface	10,347	Prune	966
Field crop	10,324	Fruit and vegetable cannery	940
Carrot	9,954	Cemetery: irrigated	888
Orange	9,360	Cole crop	865
Plum	9,293	Miscellaneous subtropical fruit	356
Melon, squash, cucumber	9,252	Cherry	347
Onion and garlic	9,217	Native pasture	315
Idle-new lands prepared for crops	9,131	Oil refinery	296
Truck, nursery, and berry crop	9,052	Cabbage	232
Walnut	8,404	Pear	219
Extractive industry	7,628	Commercial	135
Sudan	7,015	Miscellaneous establishment	122
Corn	6,633	Idle	119
Feed lot	6,559	Manufacturing	100
Residence: 3-4 houses/ac	6,096	Bushberry	88
Miscellaneous deciduous	5,970	School	88
Bean (green)	5,576	Urban: residential	88

Table D2 continued

Land use class	NDVI values (*N*)	Land use class	NDVI values (*N*)
Flowers and nursery	5,493	Pea	83
Mixed pasture	5,164	Avocado	61
Peach and nectarine	5,152	Paved area	53
Lawn area: irrigated	4,801	Motel	52
Pepper	4,759		

[1] Data from CDWR (2013)

[2] There are 645,127 NDVI values for each month between January and October 1990. There is a total of 6,451,270 NDVI values across all months contributing to the crop signature library.

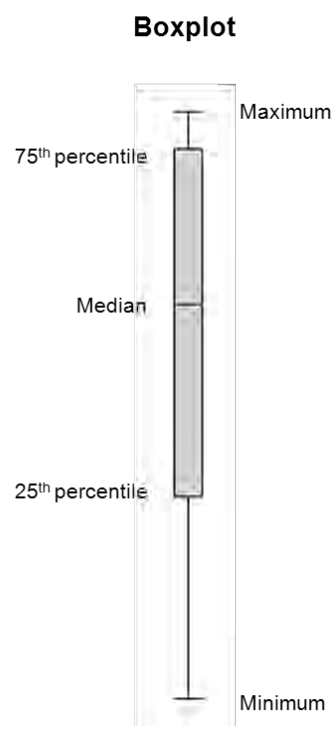


Figure D1 Boxplot characteristics

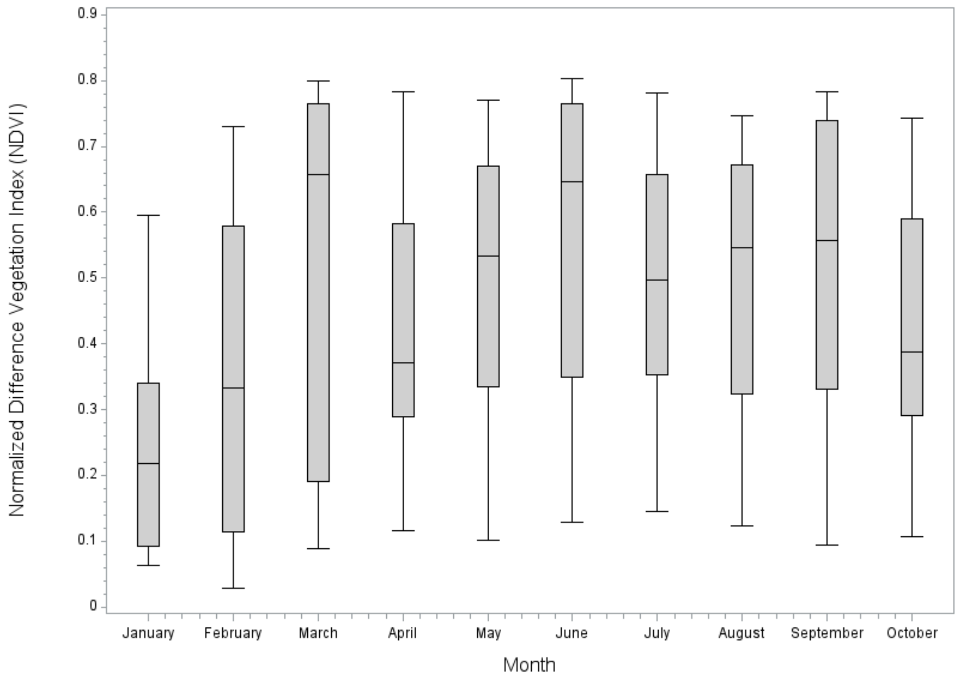


Figure D2 Alfalfa: NDVI in Kern County, 1990

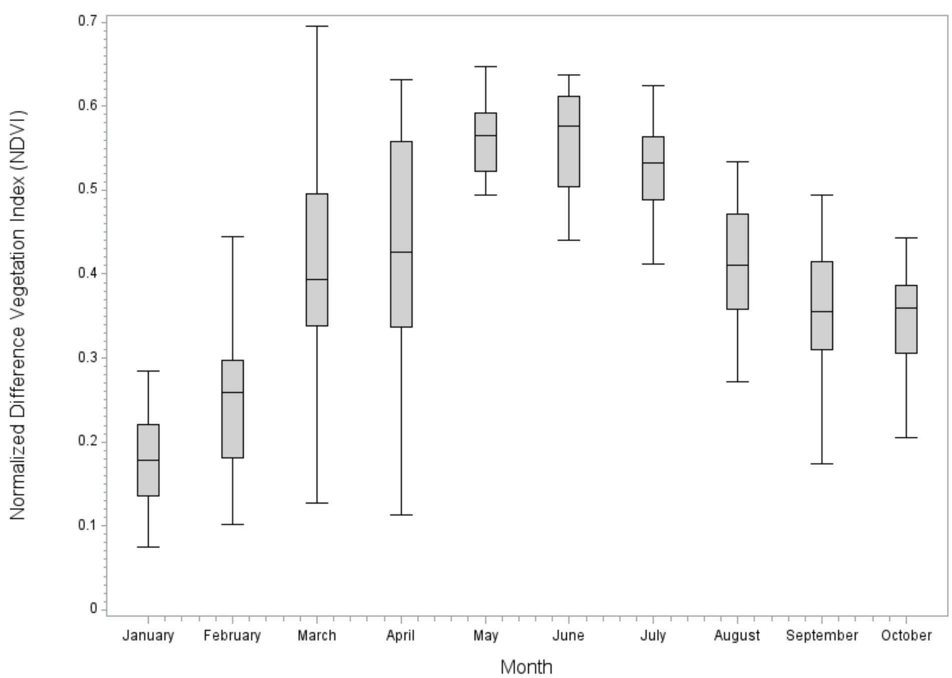


Figure D3 Almond: NDVI in Kern County, 1990

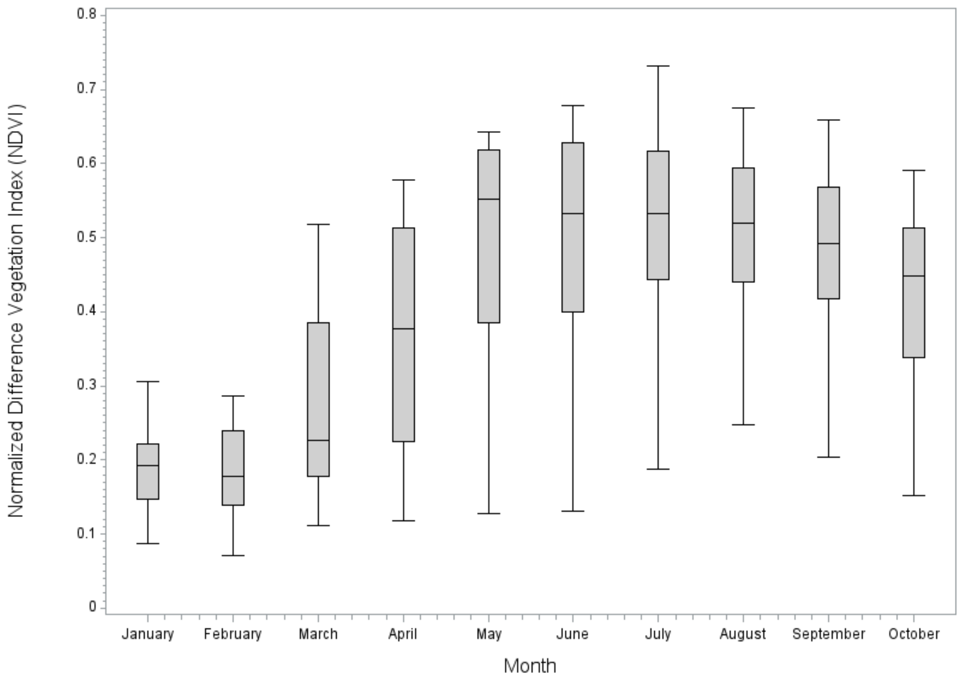


Figure D4 Apple: NDVI in Kern County, 1990

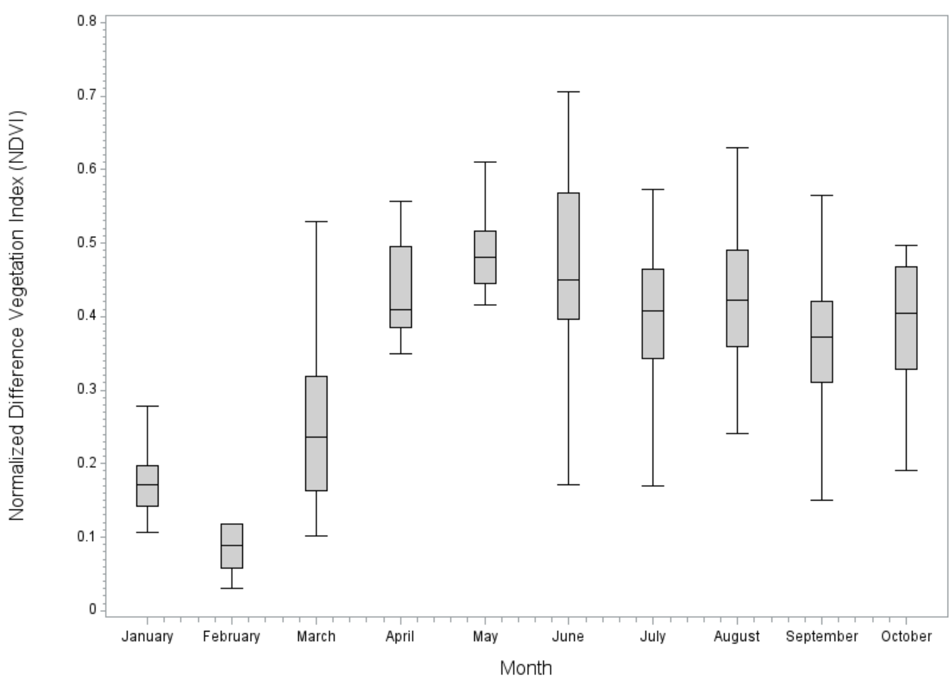


Figure D5 Apricot: NDVI in Kern County, 1990

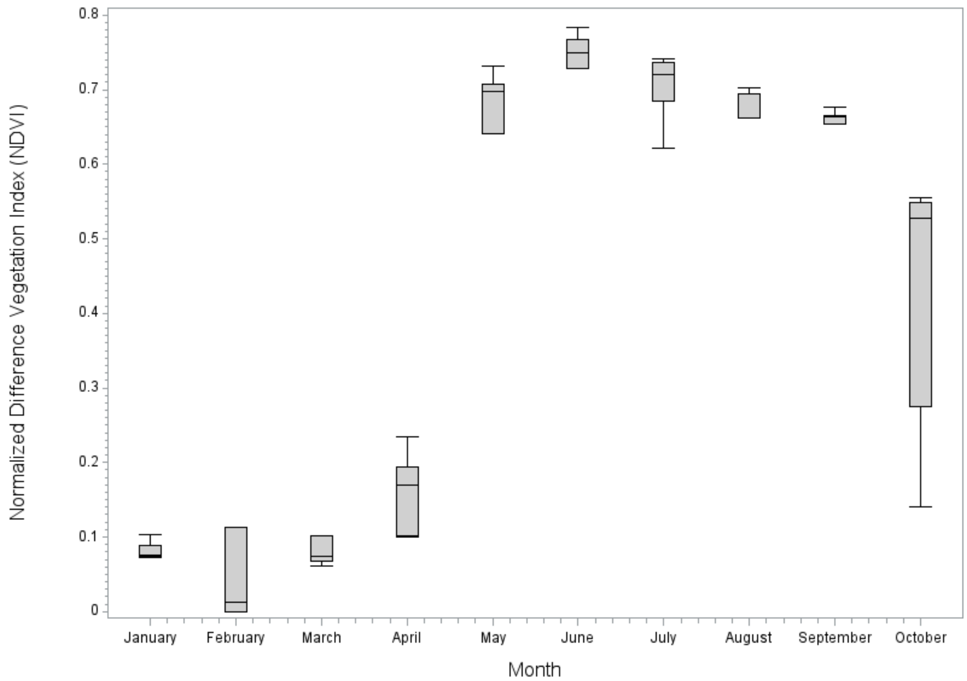


Figure D6 Asparagus: NDVI in Kern County, 1990

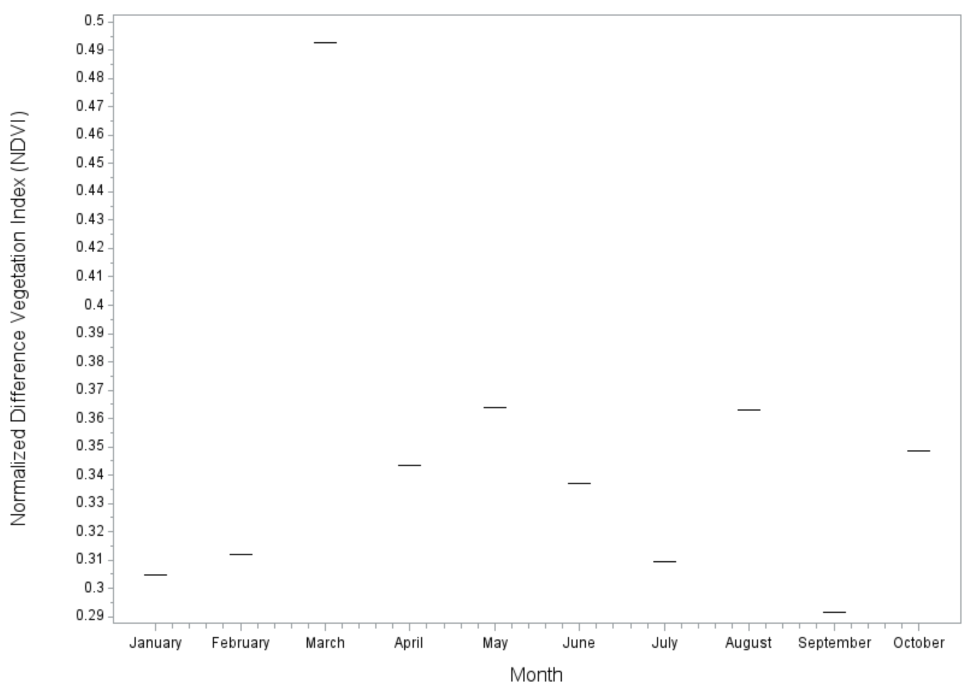


Figure D7 Avocado: NDVI in Kern County, 1990

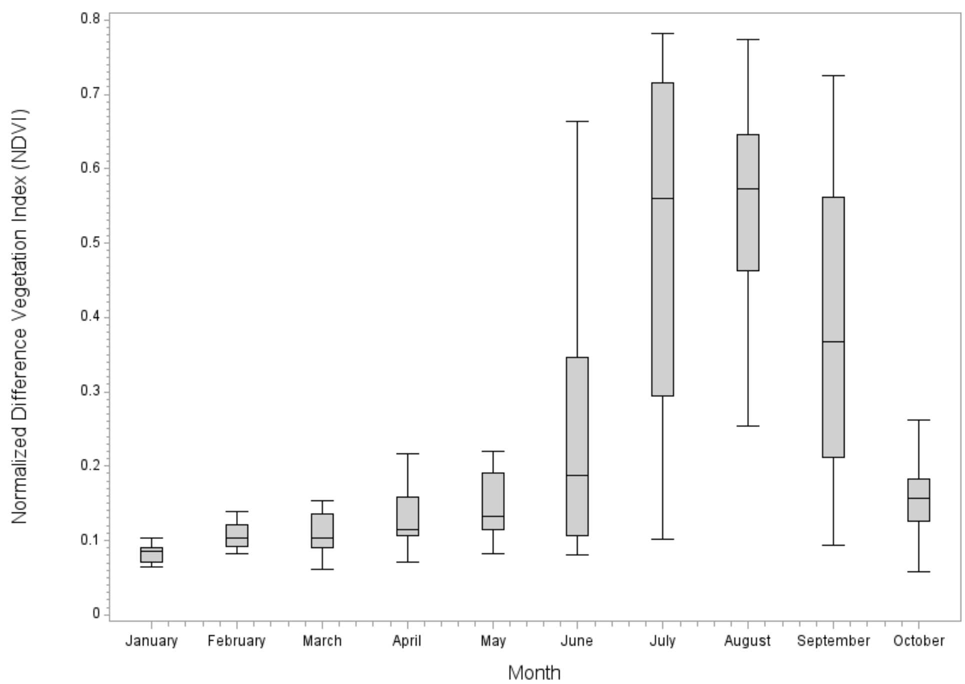


Figure D8 Bean (dry): NDVI in Kern County, 1990

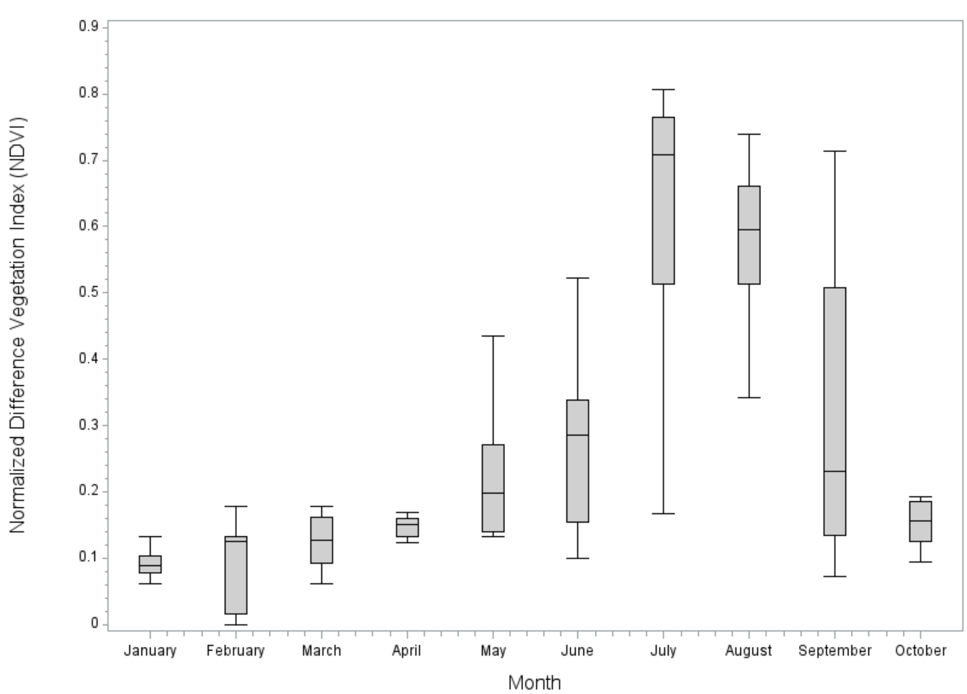


Figure D9 Bean (green): NDVI in Kern County, 1990

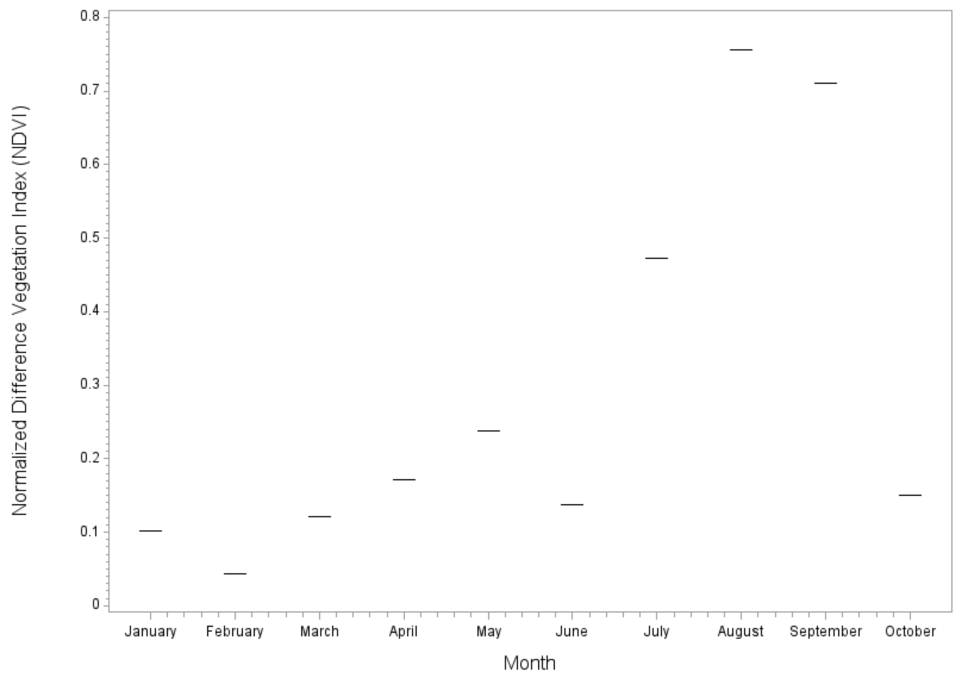


Figure D10 Bushberry: NDVI in Kern County, 1990

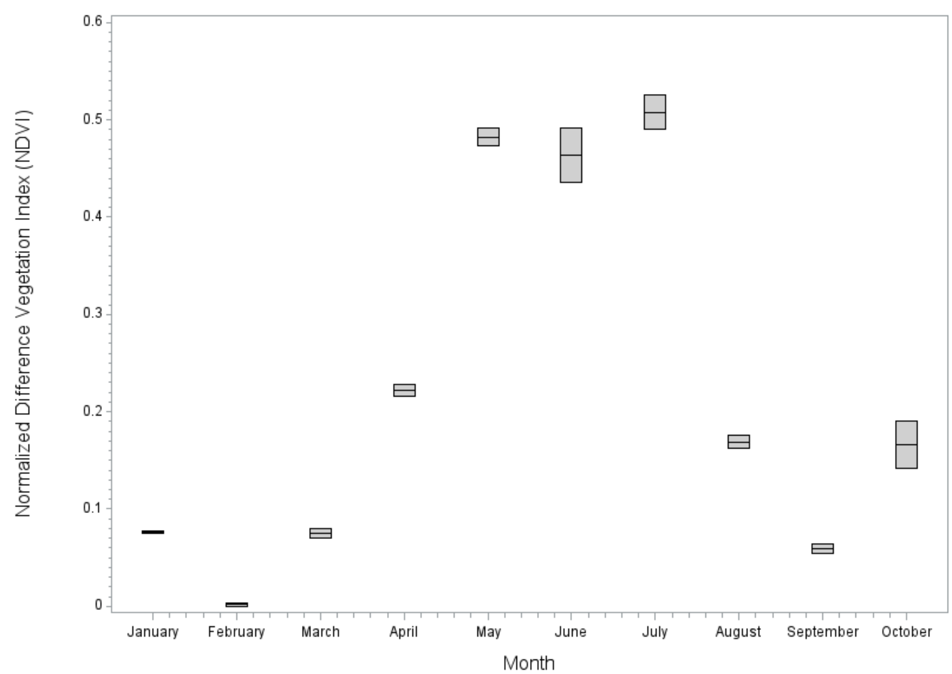


Figure D11 Cabbage: NDVI in Kern County, 1990

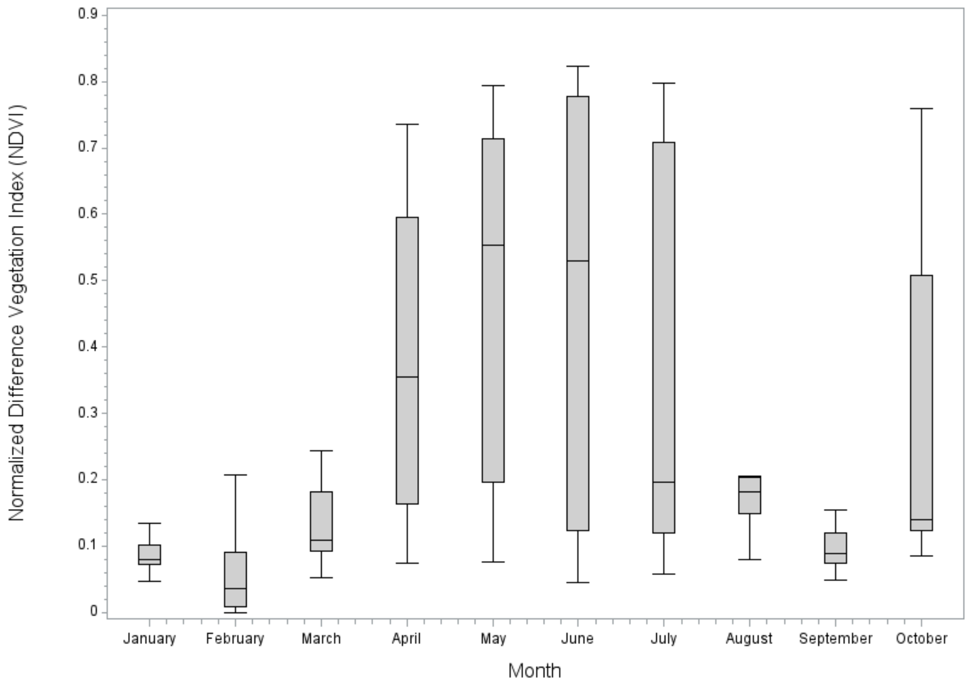


Figure D12 Carrot: NDVI in Kern County, 1990

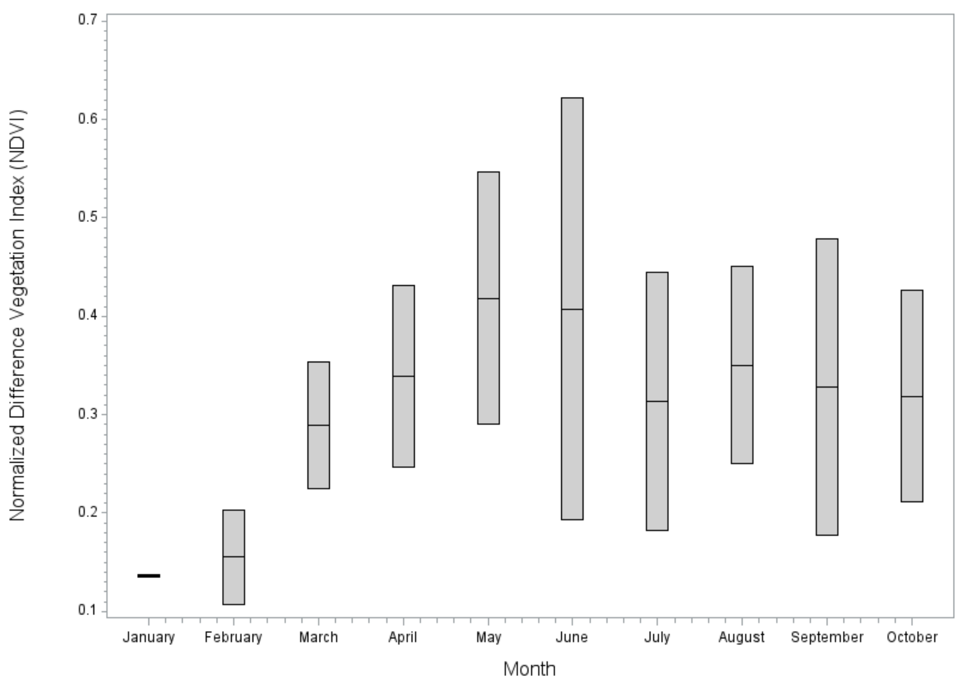


Figure D13 Cherry: NDVI in Kern County, 1990

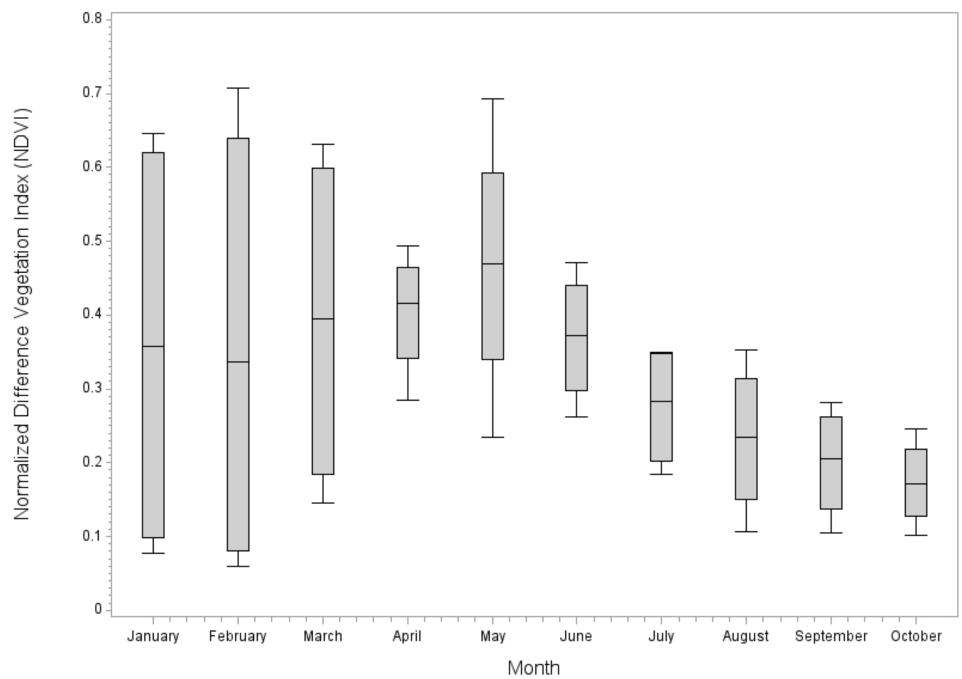


Figure D14 Cole crop: NDVI in Kern County, 1990

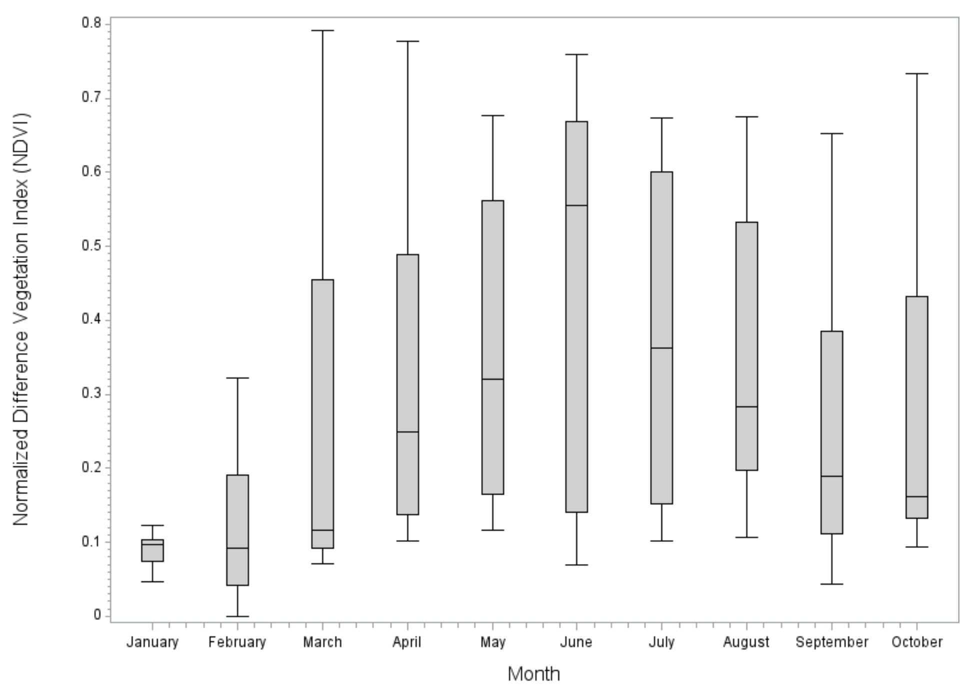


Figure D15 Corn: NDVI in Kern County, 1990

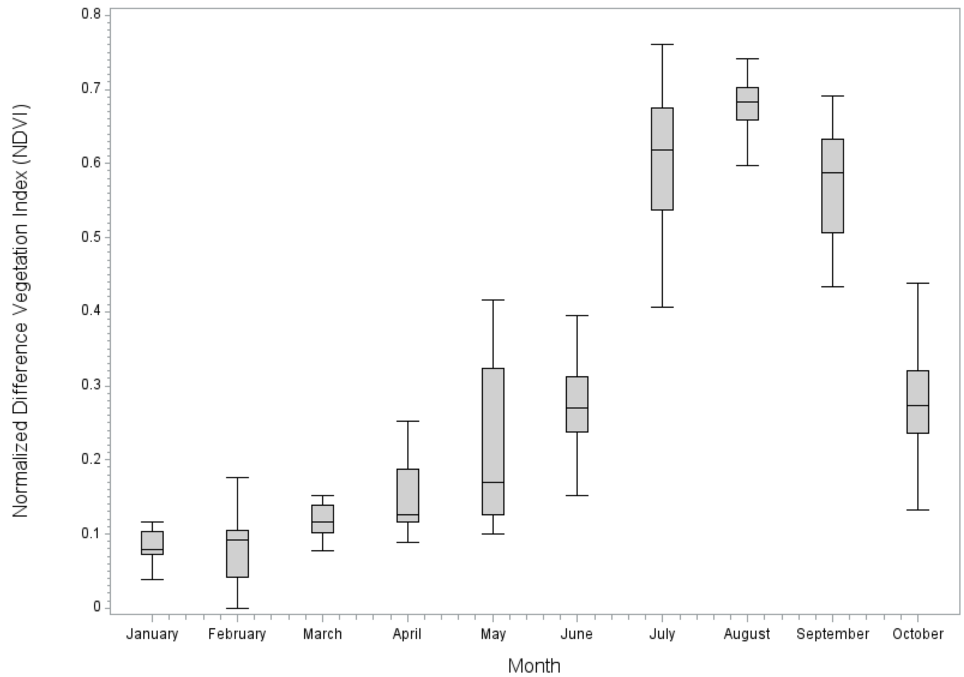


Figure D16 Cotton: NDVI in Kern County, 1990

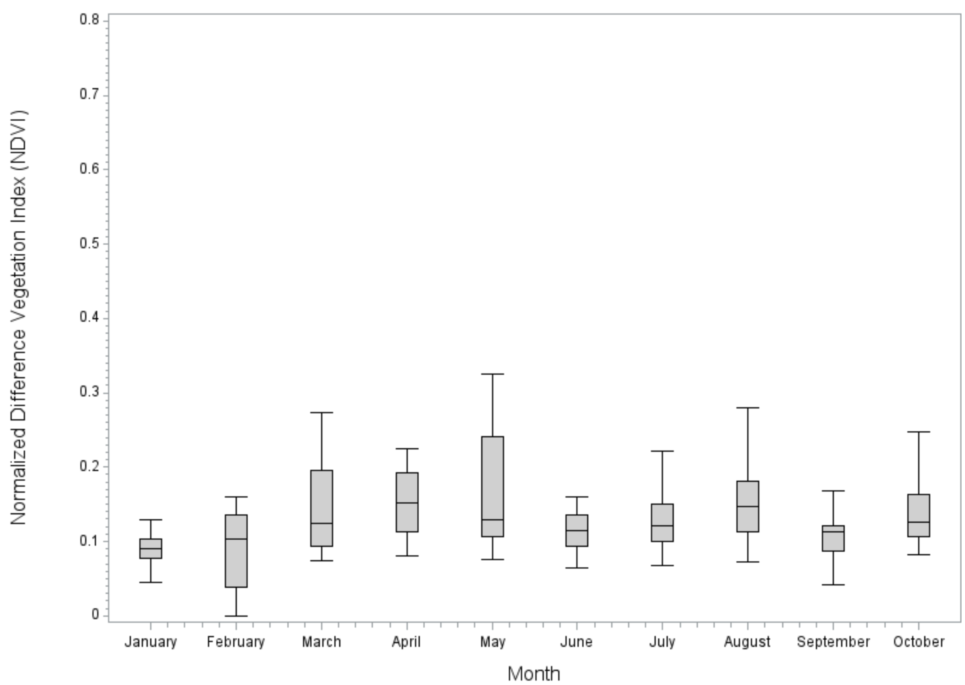


Figure D17 Field crop: NDVI in Kern County, 1990

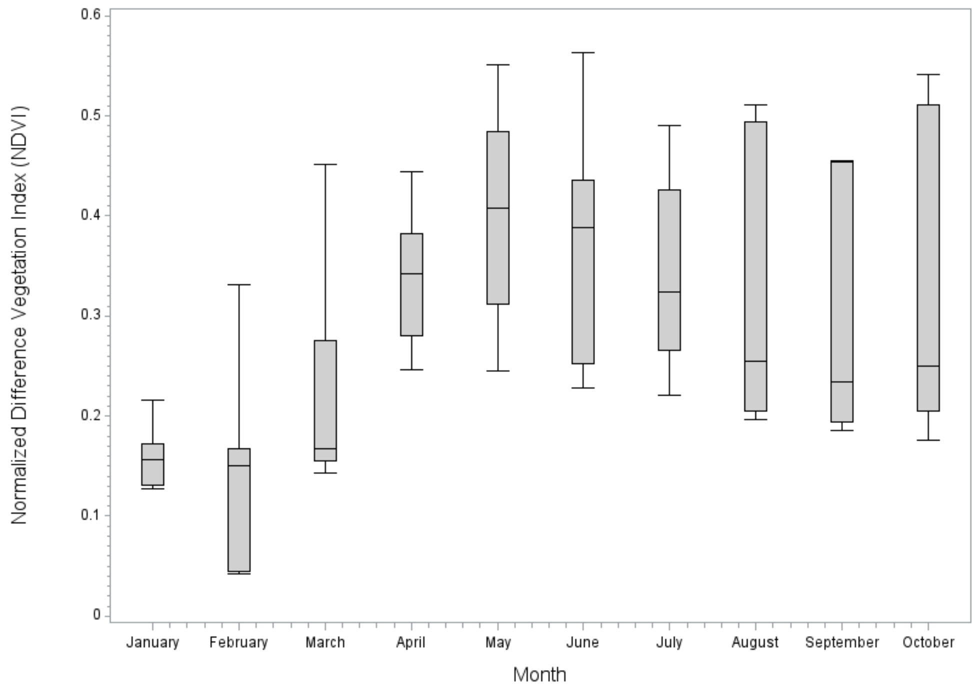


Figure D18 Fig: NDVI in Kern County, 1990

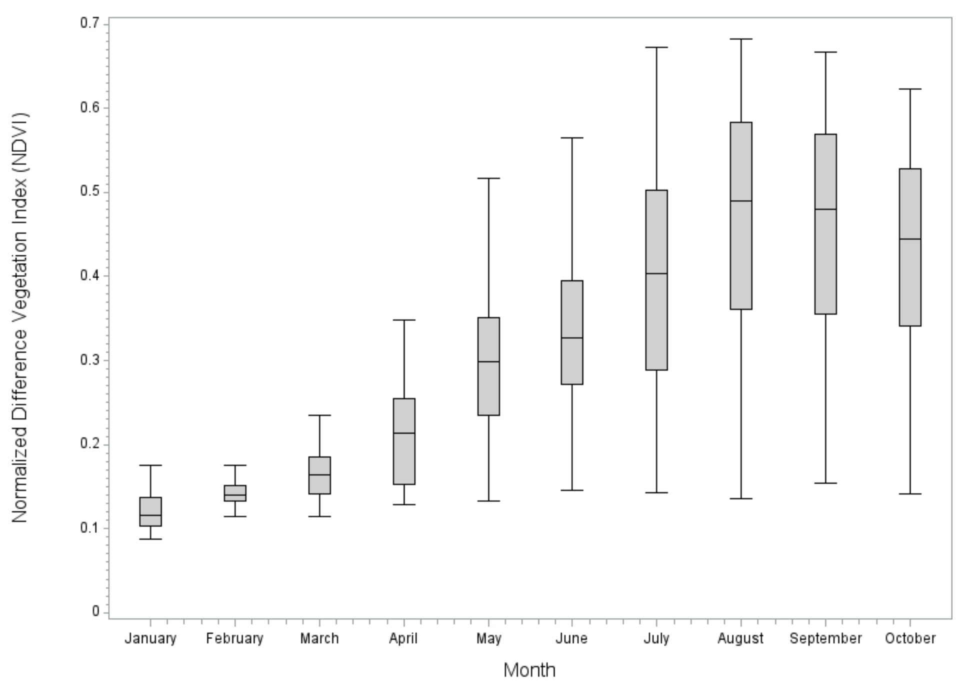


Figure D19 Flowers and nursery: NDVI in Kern County, 1990

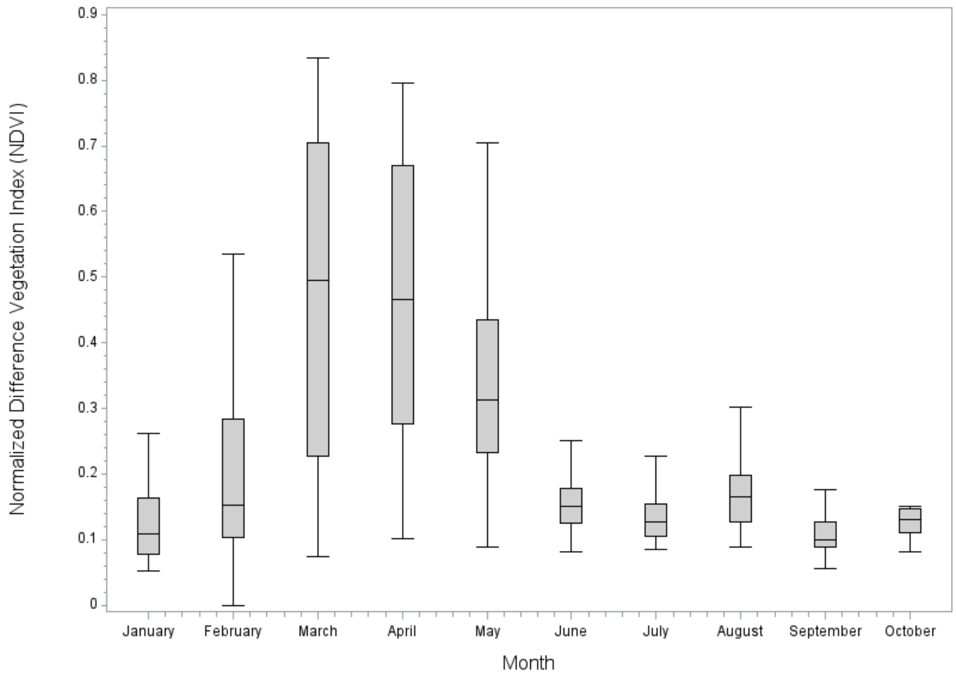


Figure D20 Grain and hay crop: NDVI in Kern County, 1990

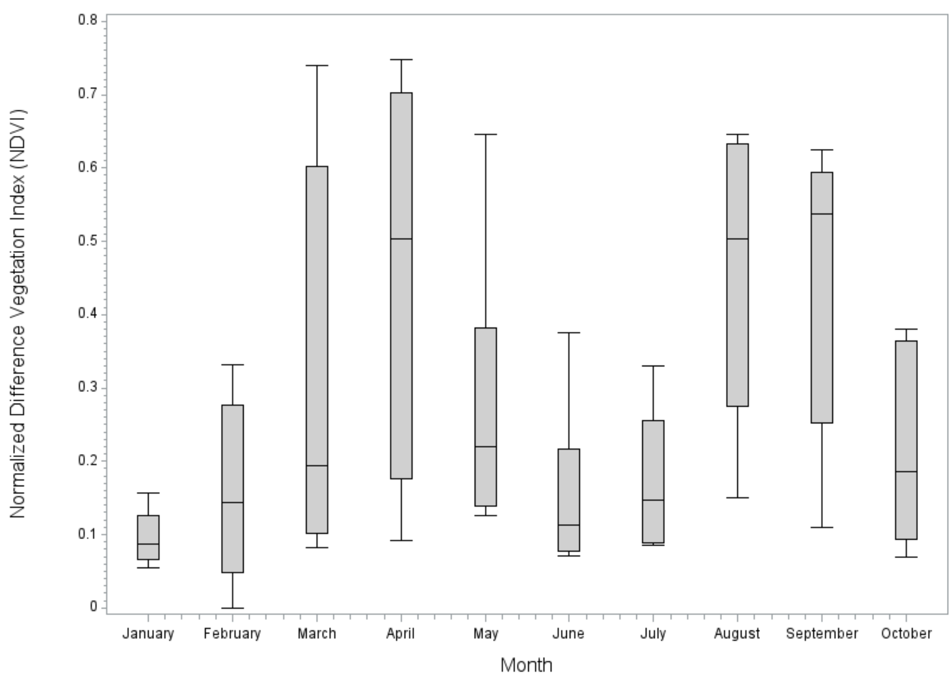


Figure D21 Grain sorghum: NDVI in Kern County, 1990

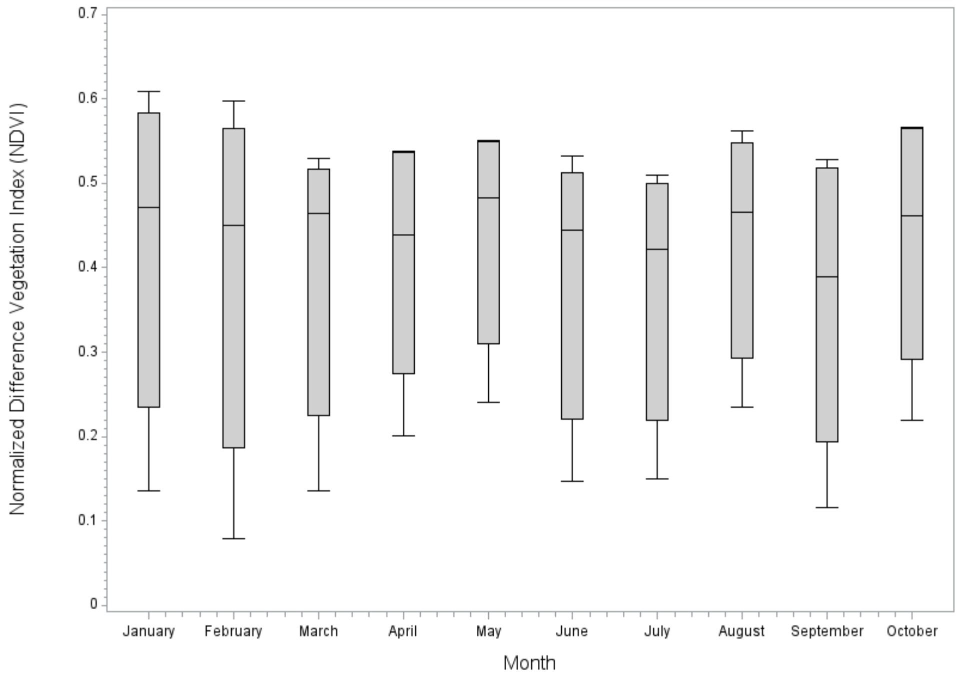


Figure D22 Grapefruit: NDVI in Kern County, 1990

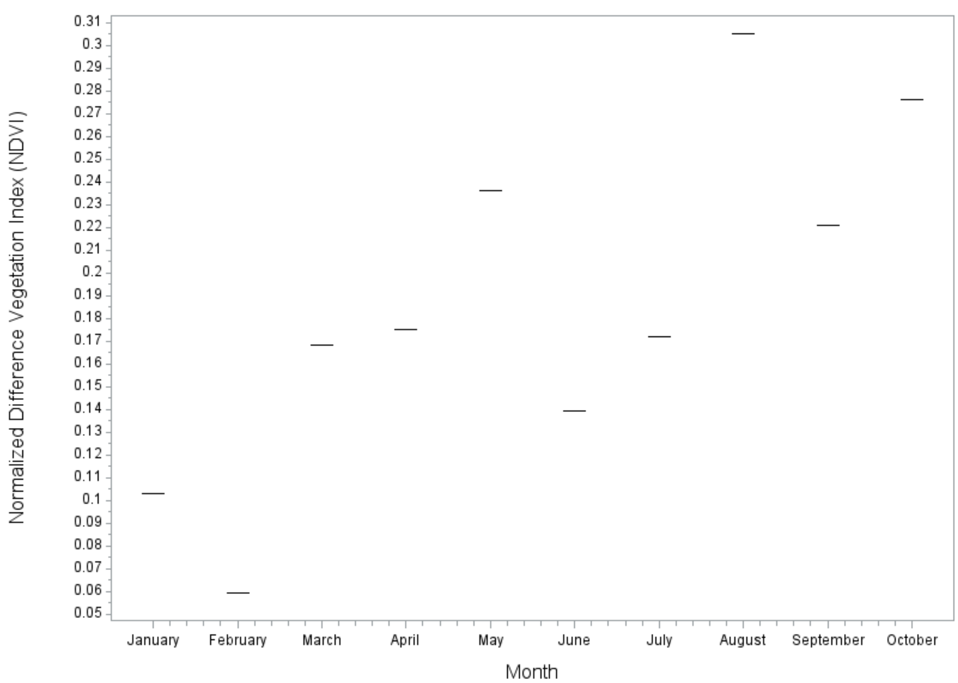


Figure D23 Idle: NDVI in Kern County, 1990

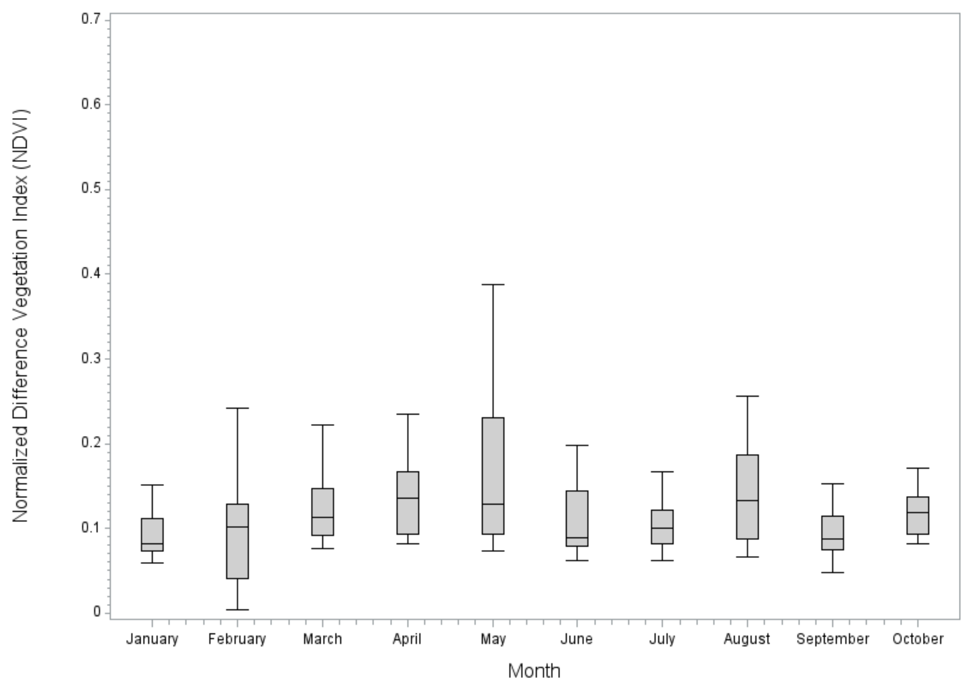


Figure D24 Idle-cropped in past year: NDVI in Kern County, 1990

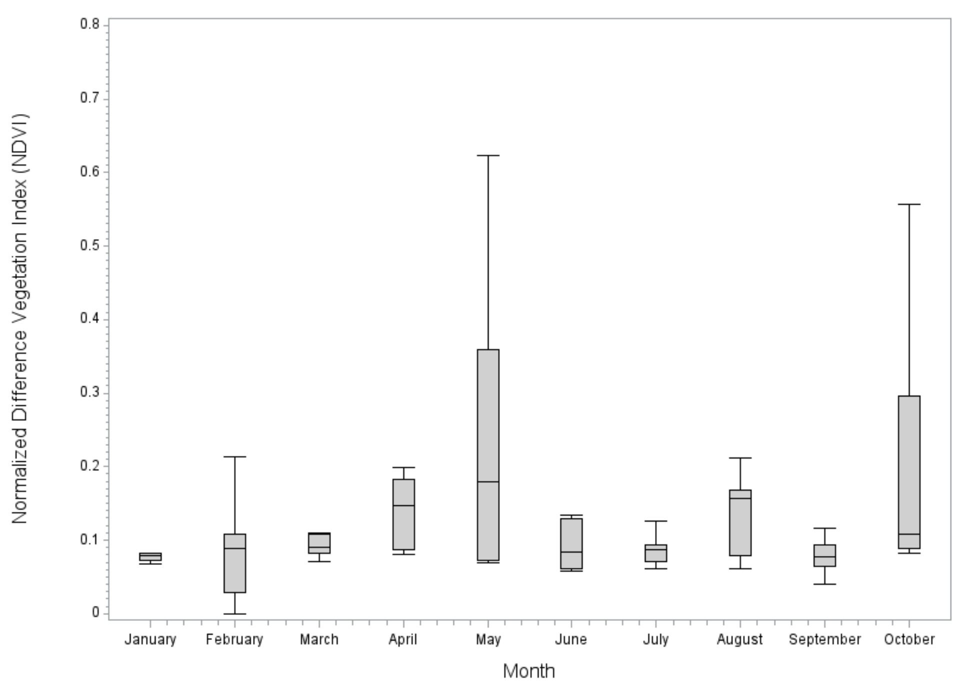


Figure D25 Idle-new lands prepared for crops: NDVI in Kern County, 1990

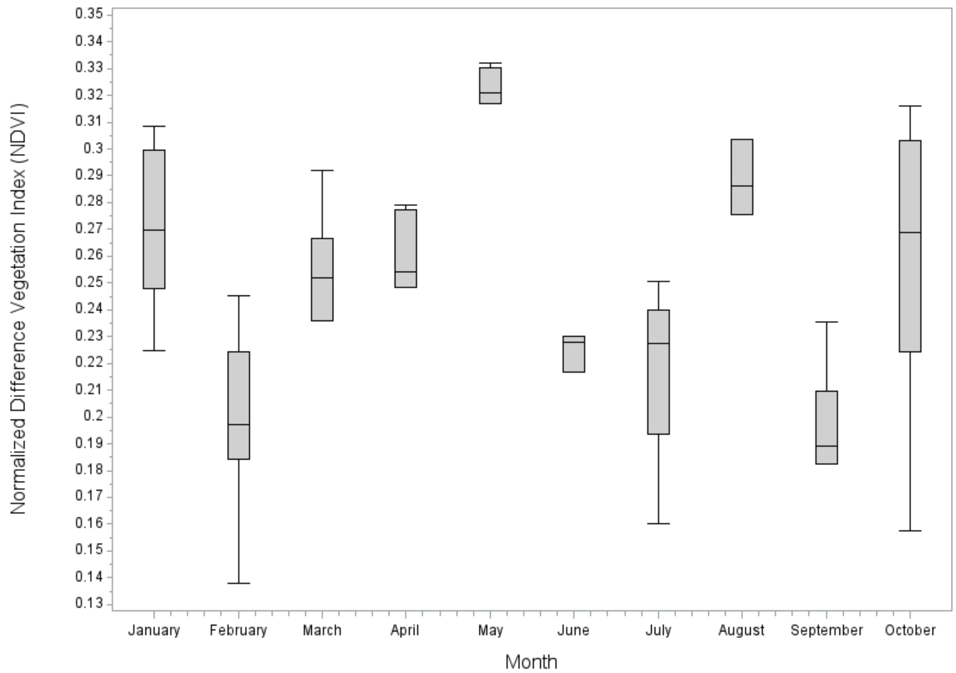


Figure D26 Jojoba: NDVI in Kern County, 1990

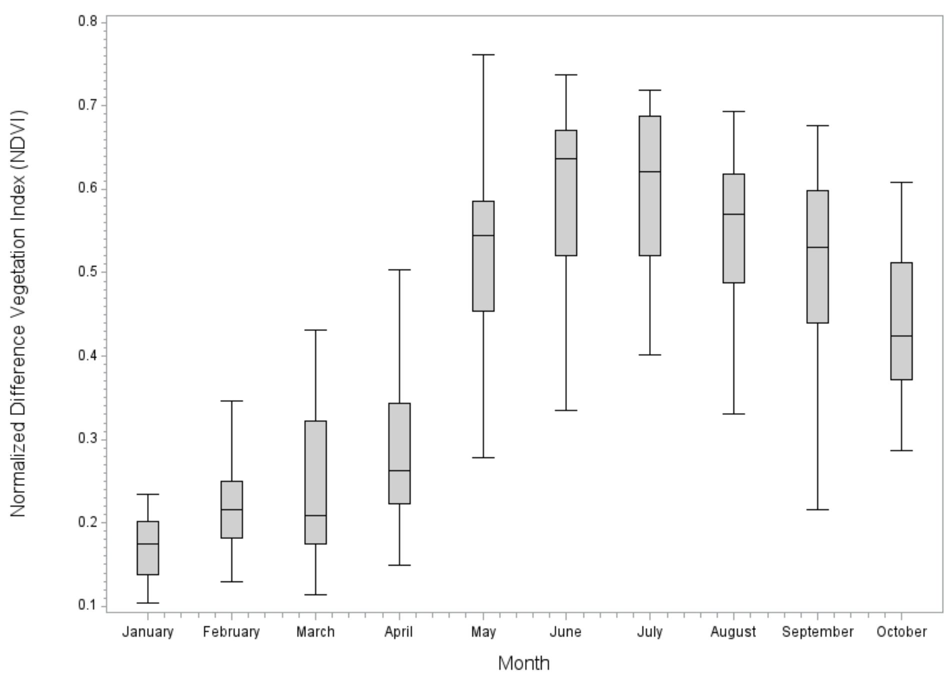


Figure D27 Kiwi: NDVI in Kern County, 1990

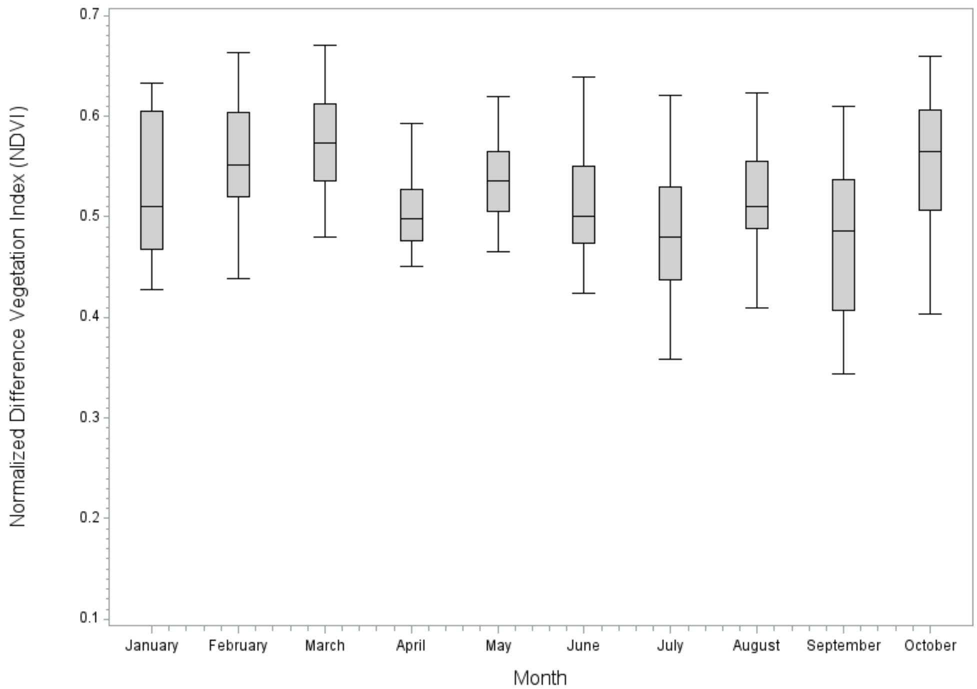


Figure D28 Lemon: NDVI in Kern County, 1990

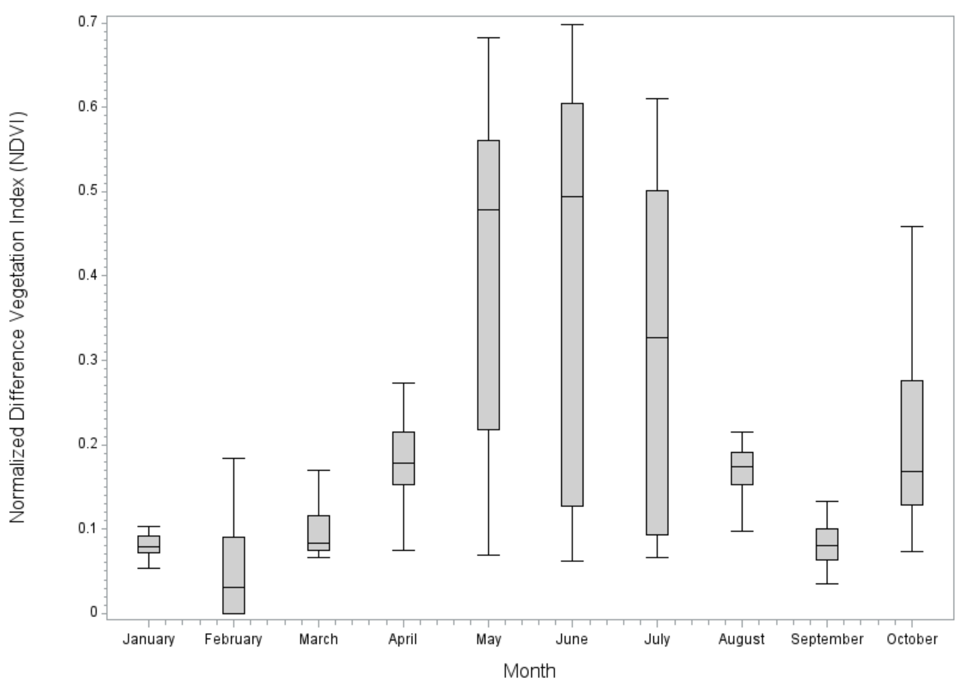


Figure D29 Lettuce: NDVI in Kern County, 1990

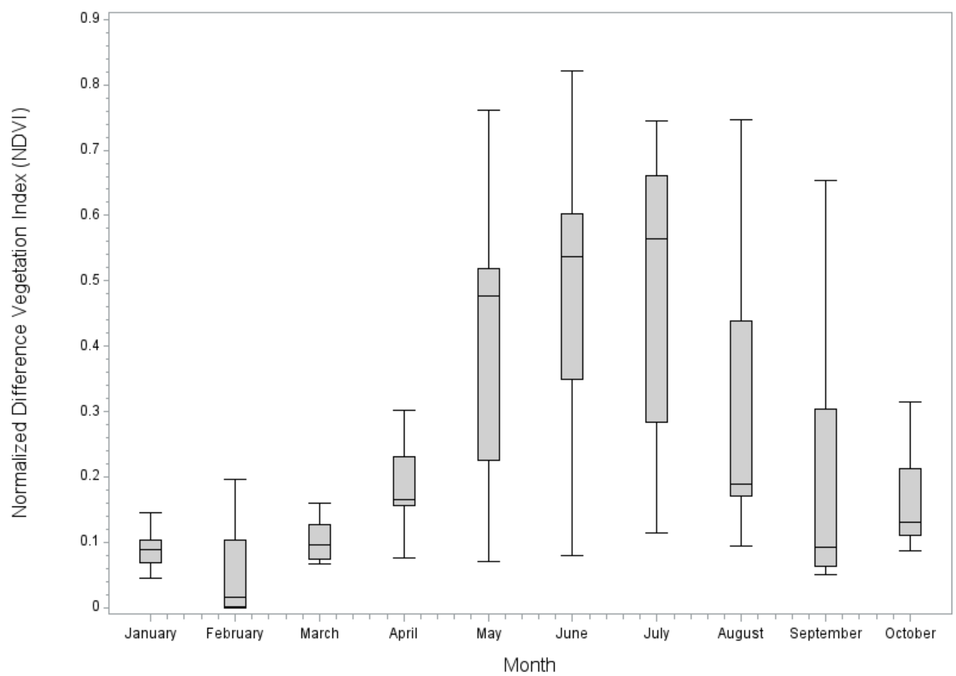


Figure D30 Melon, squash, cucumber: NDVI in Kern County, 1990

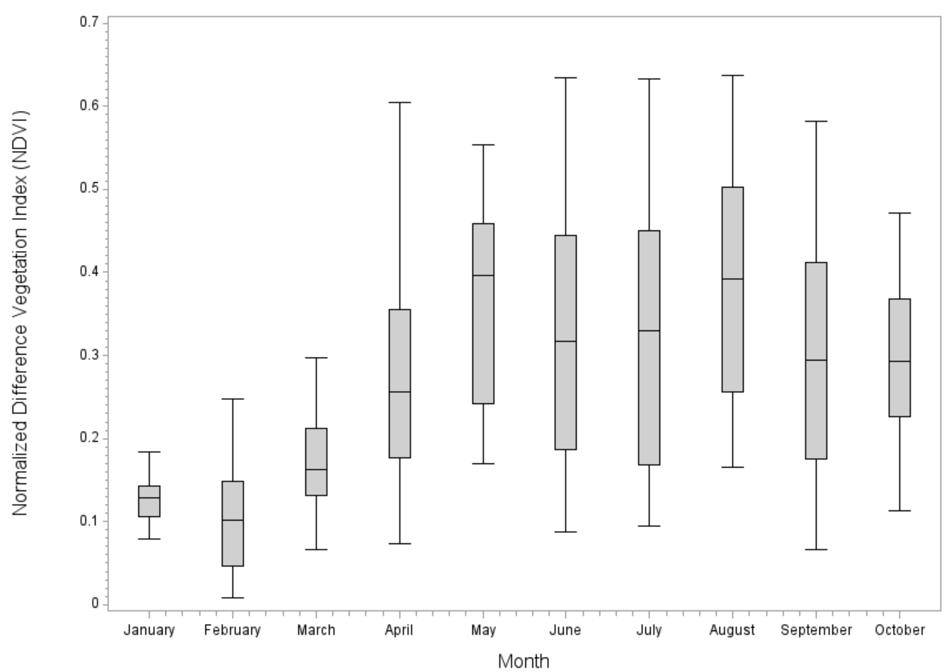


Figure D31 Miscellaneous deciduous: NDVI in Kern County, 1990

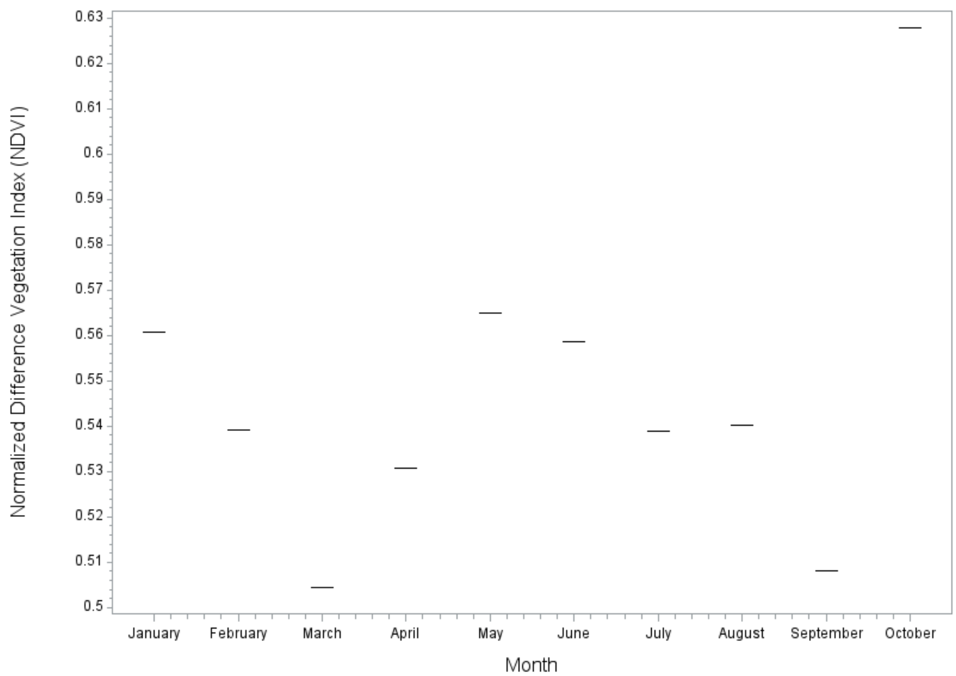


Figure D32 Miscellaneous subtropical fruit: NDVI in Kern County, 1990

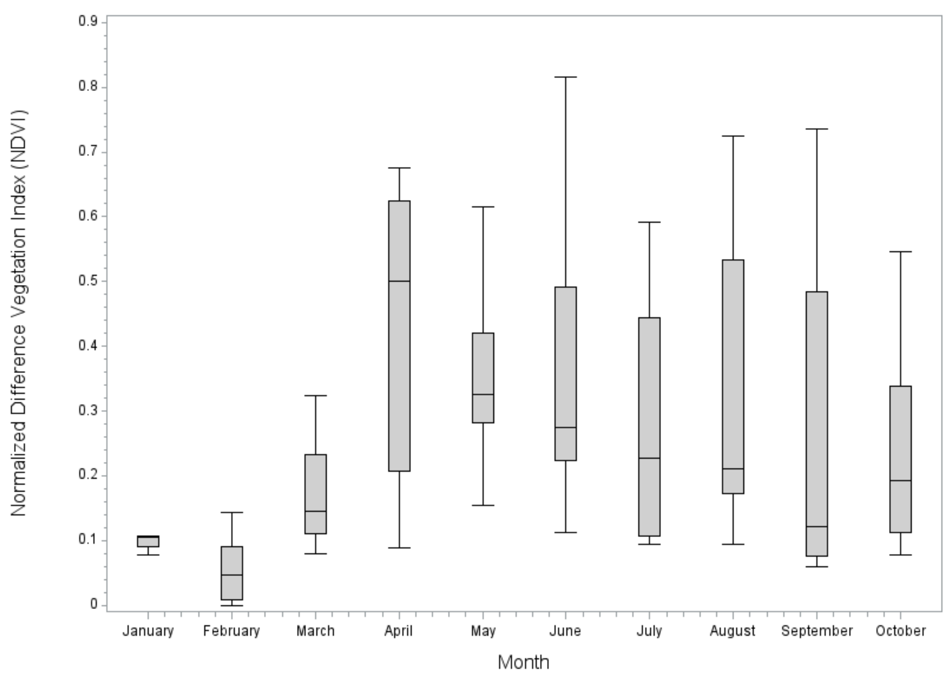


Figure D33 Miscellaneous truck: NDVI in Kern County, 1990

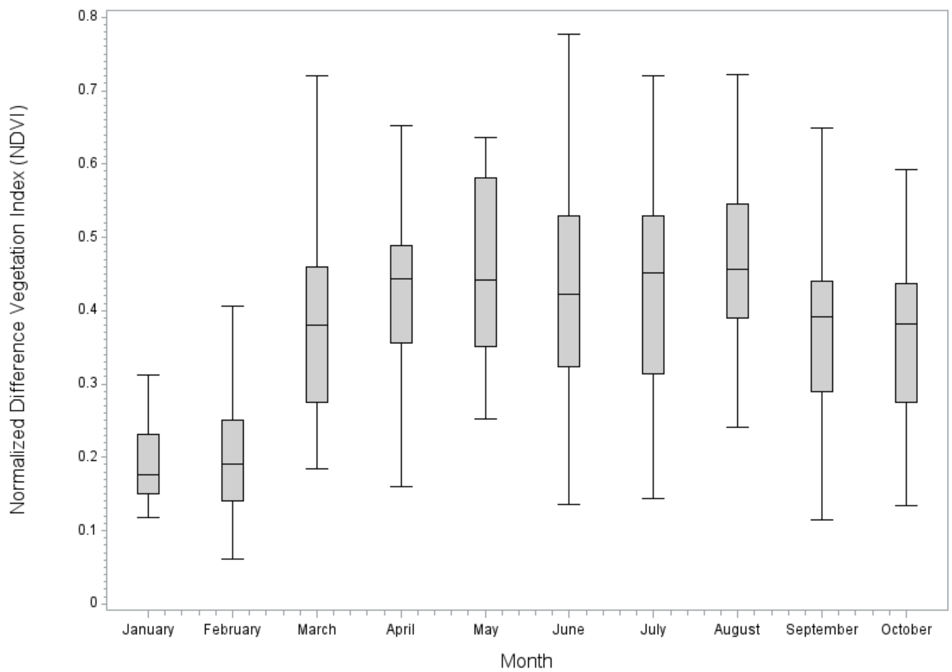


Figure D34 Mixed pasture: NDVI in Kern County, 1990

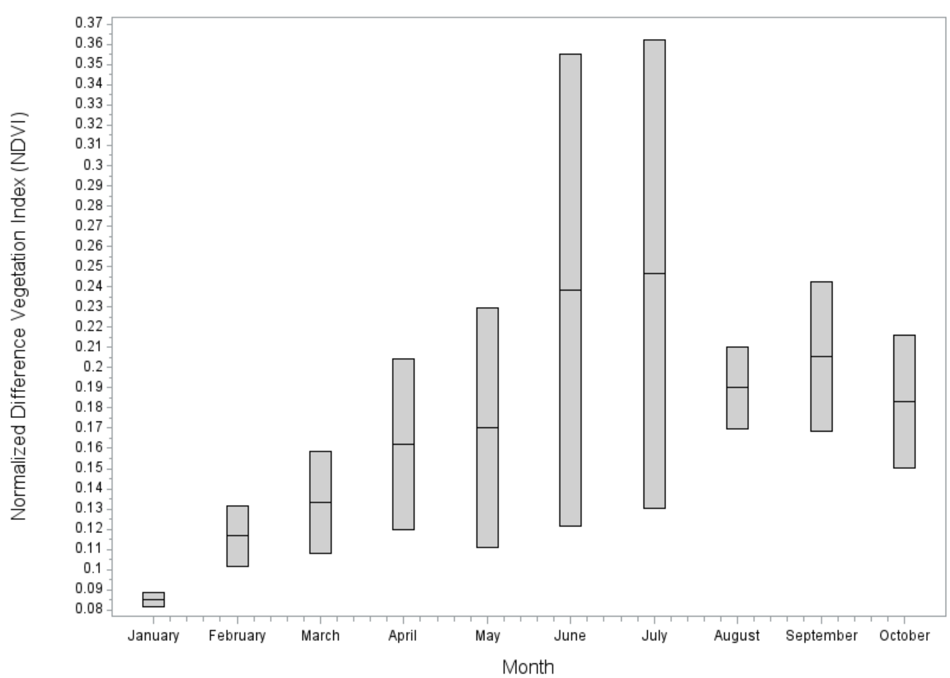


Figure D35 Native pasture: NDVI in Kern County, 1990

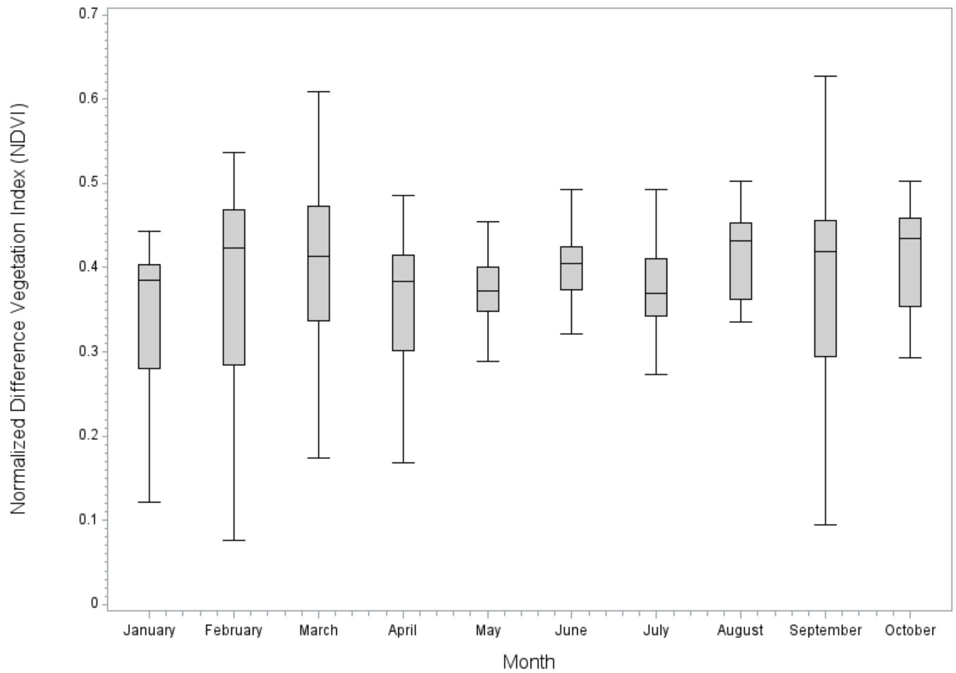


Figure D36 Olive: NDVI in Kern County, 1990

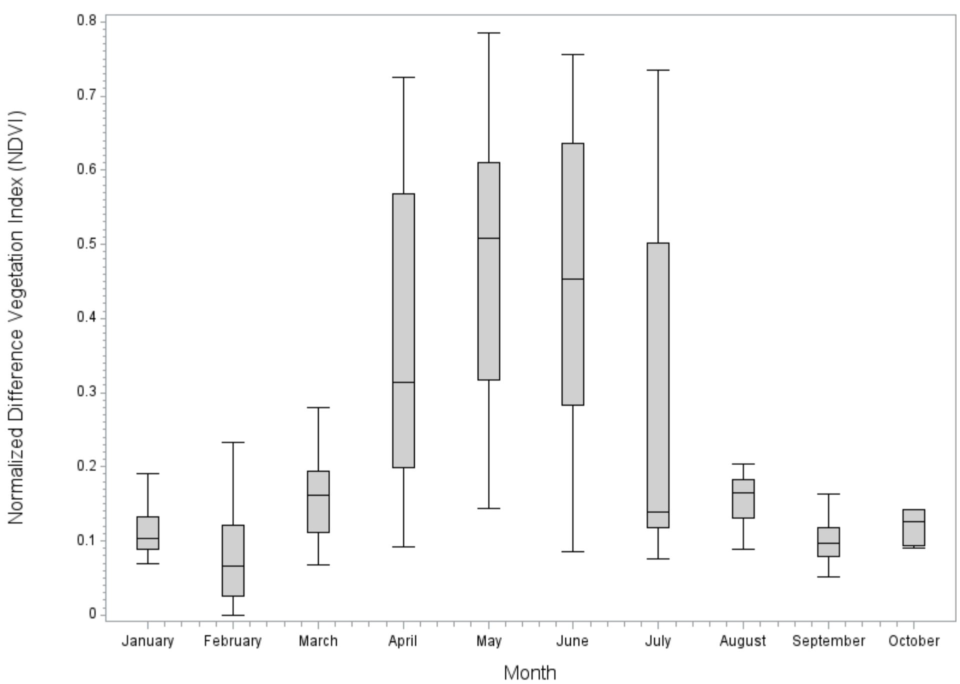


Figure D37 Onion and garlic: NDVI in Kern County, 1990

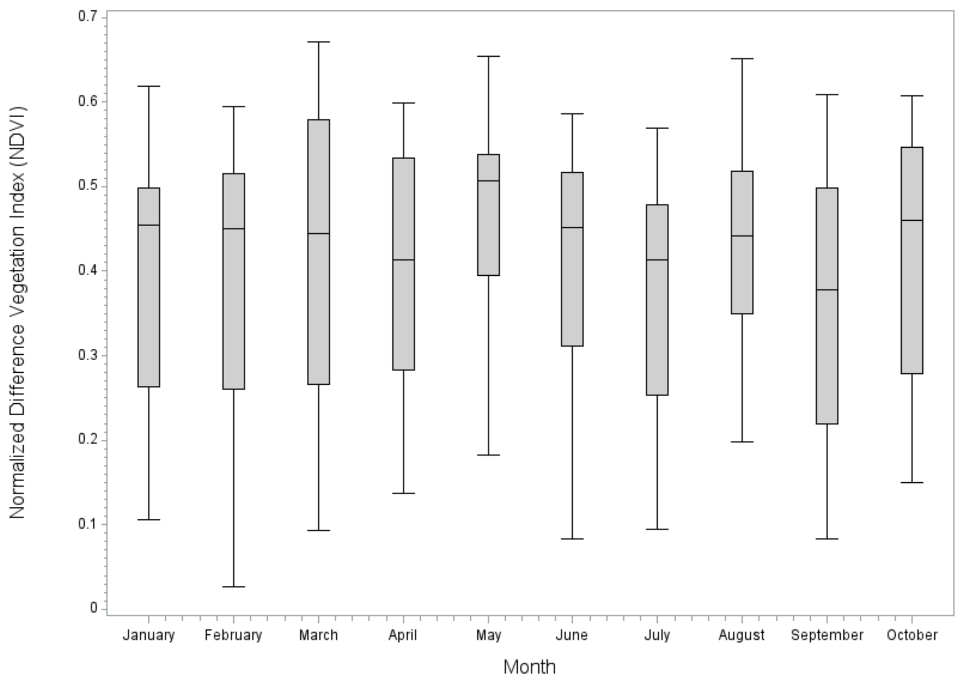


Figure D38 Orange: NDVI in Kern County, 1990

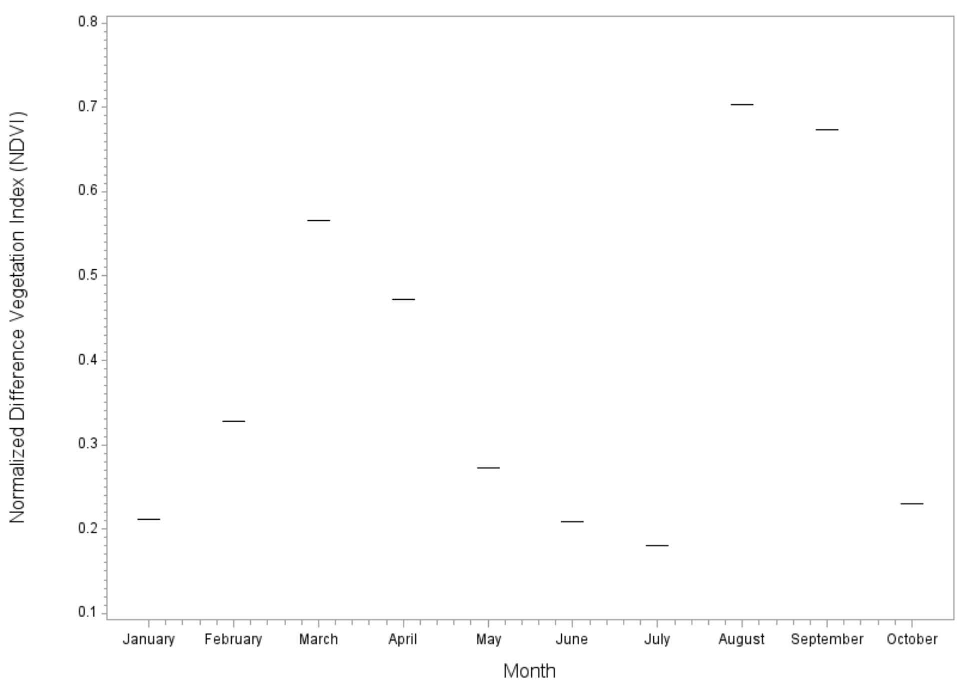


Figure D39 Pea: NDVI in Kern County, 1990

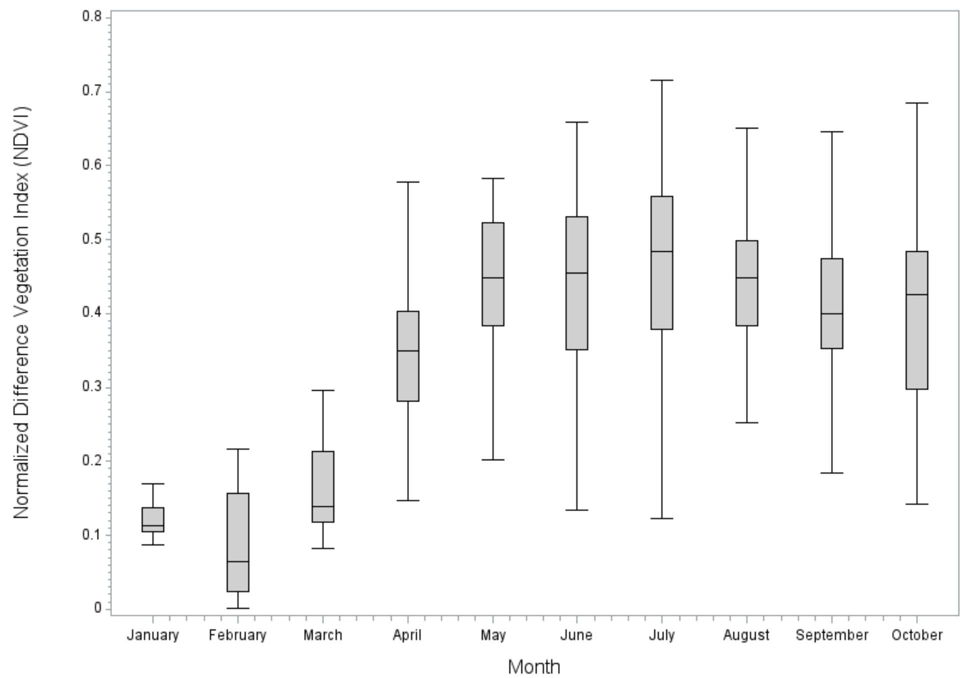


Figure D40 Peach and nectarine: NDVI in Kern County, 1990

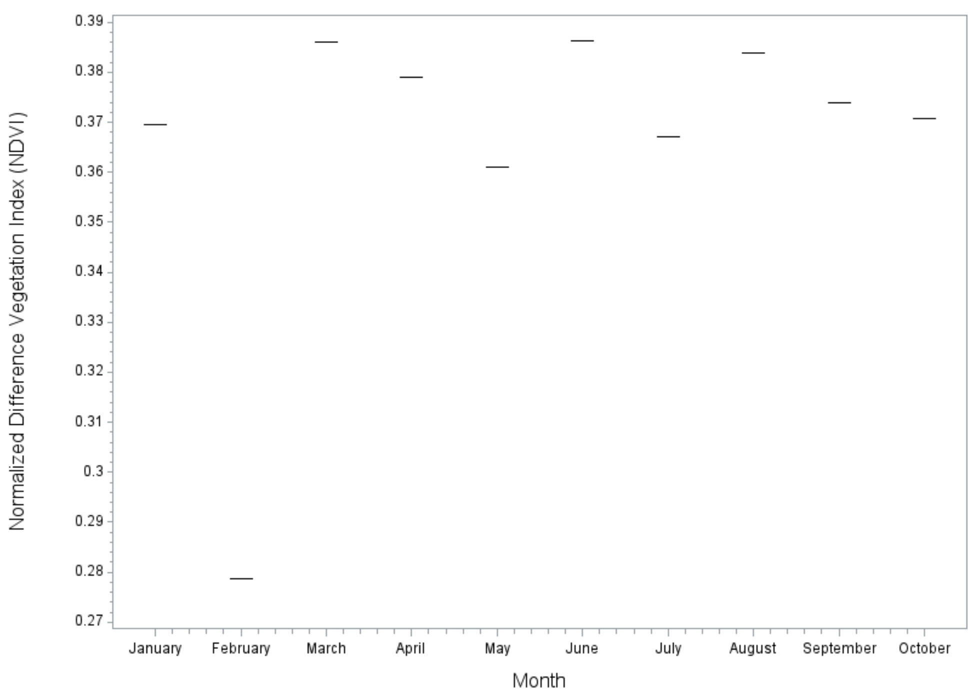


Figure D41 Pear: NDVI in Kern County, 1990

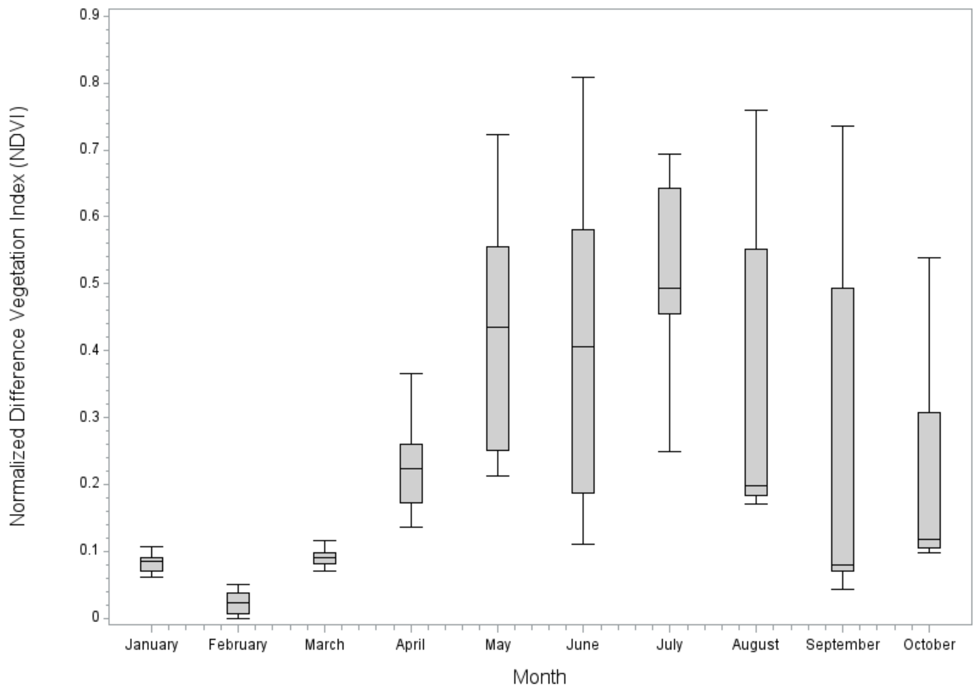


Figure D42 Pepper: NDVI in Kern County, 1990

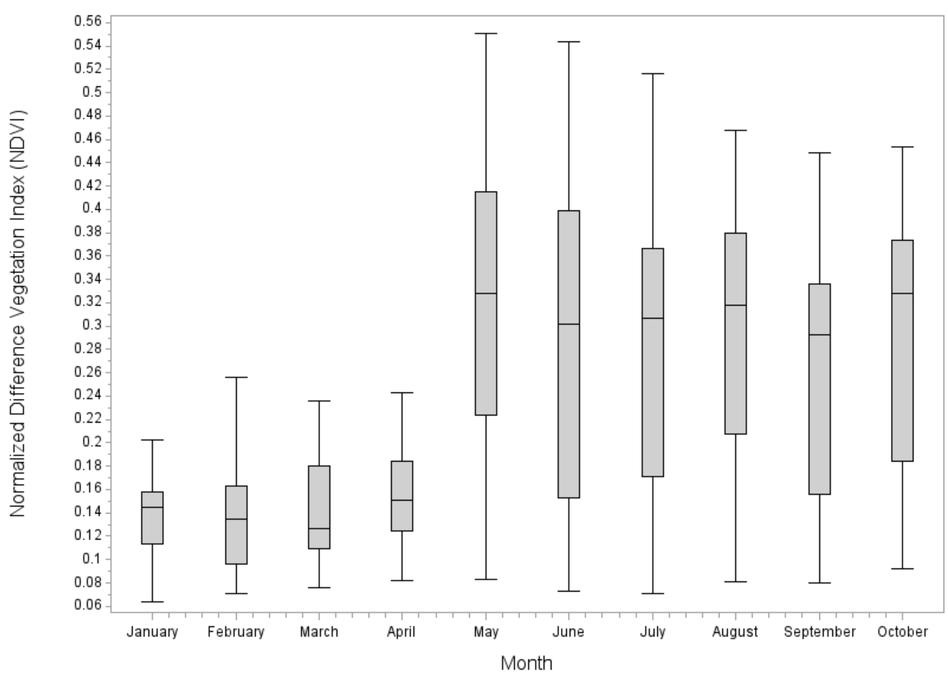


Figure D43 Pistachio: NDVI in Kern County, 1990

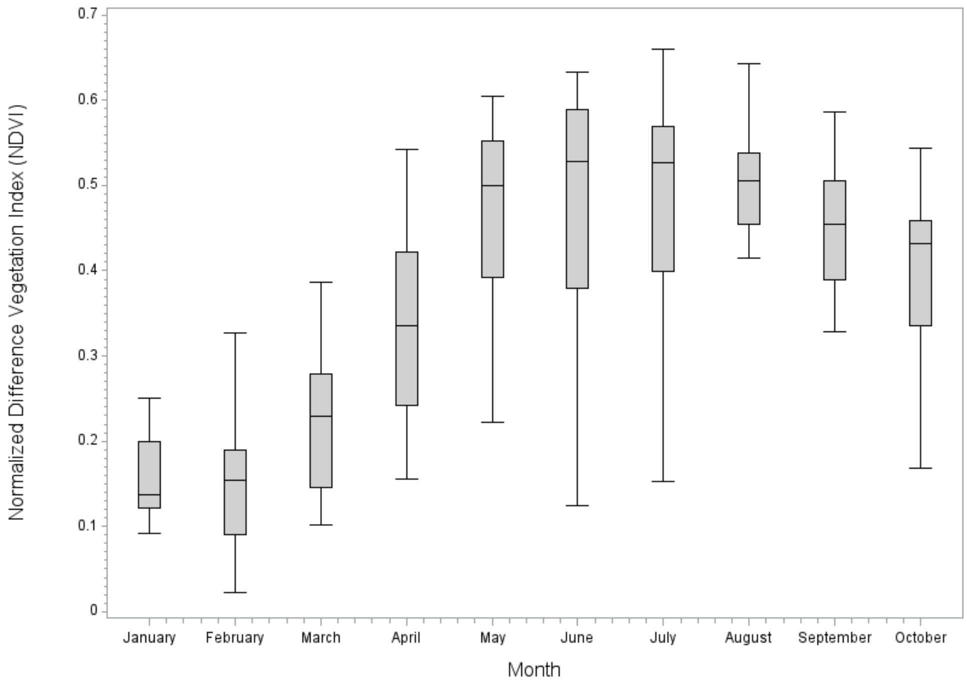


Figure D44 Plum: NDVI in Kern County, 1990

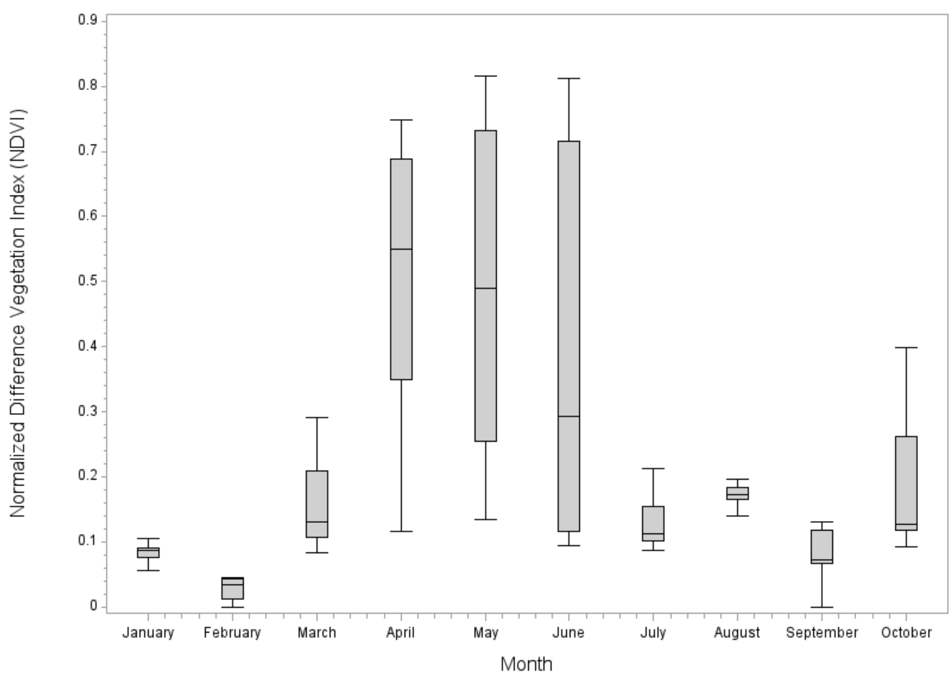


Figure D45 Potato: NDVI in Kern County, 1990

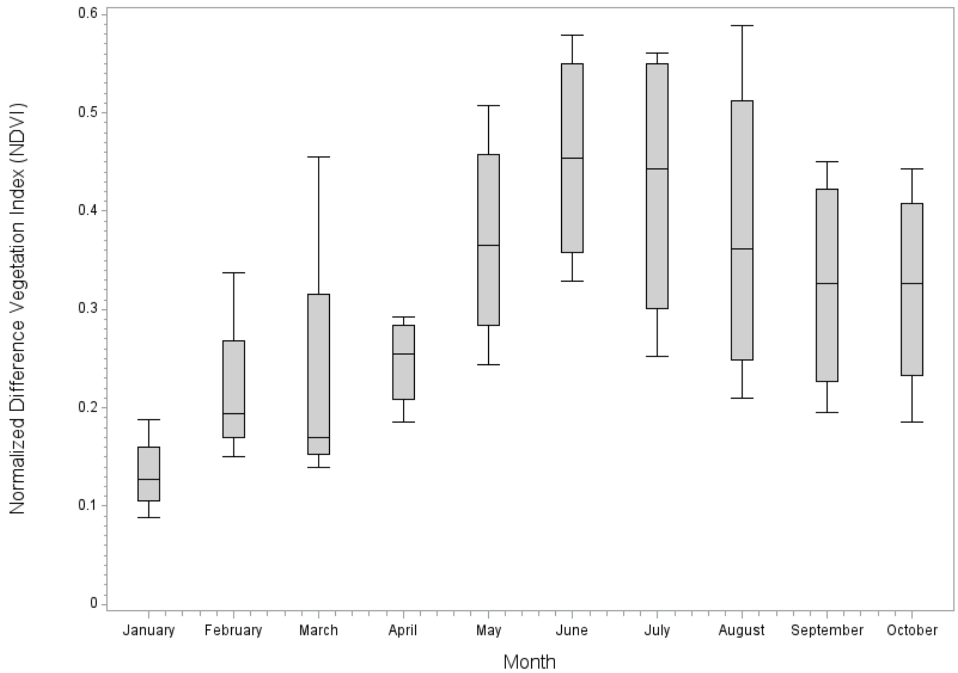


Figure D46 Prune: NDVI in Kern County, 1990

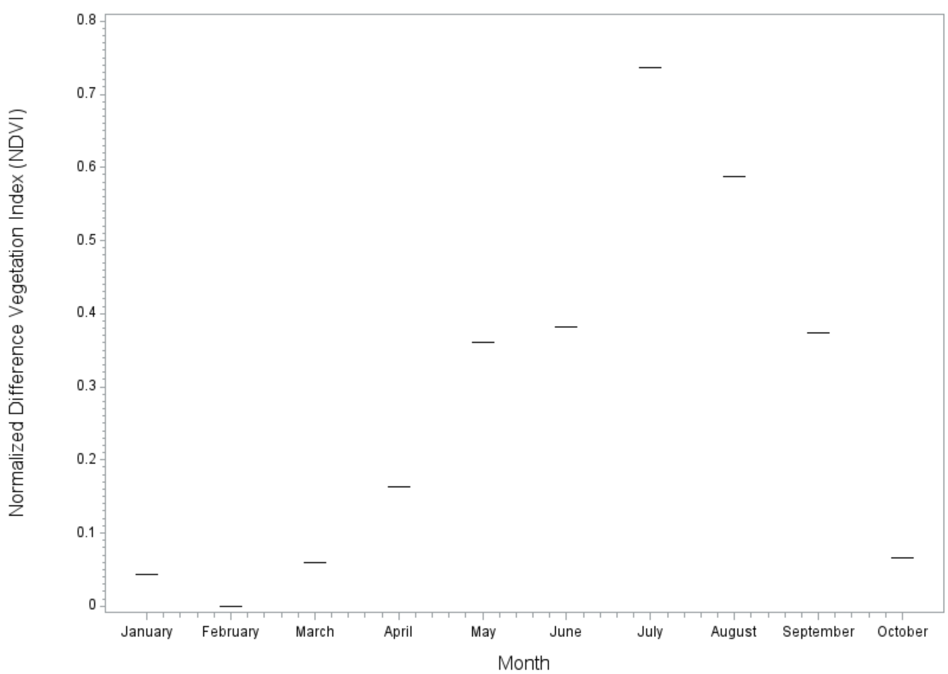


Figure D47 Rice: NDVI in Kern County, 1990

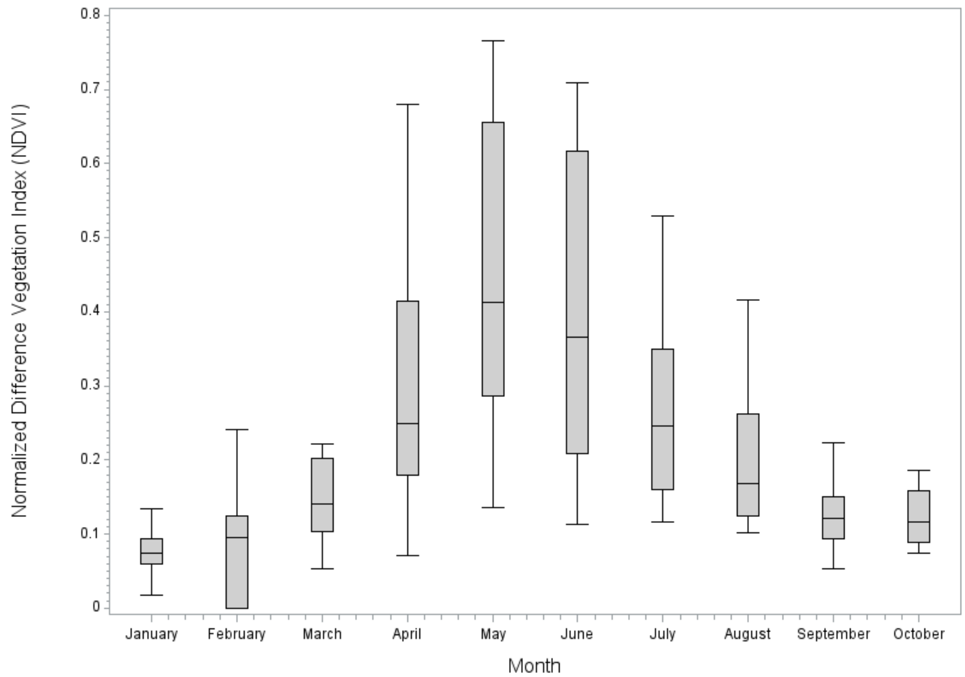


Figure D48 Safflower: NDVI in Kern County, 1990

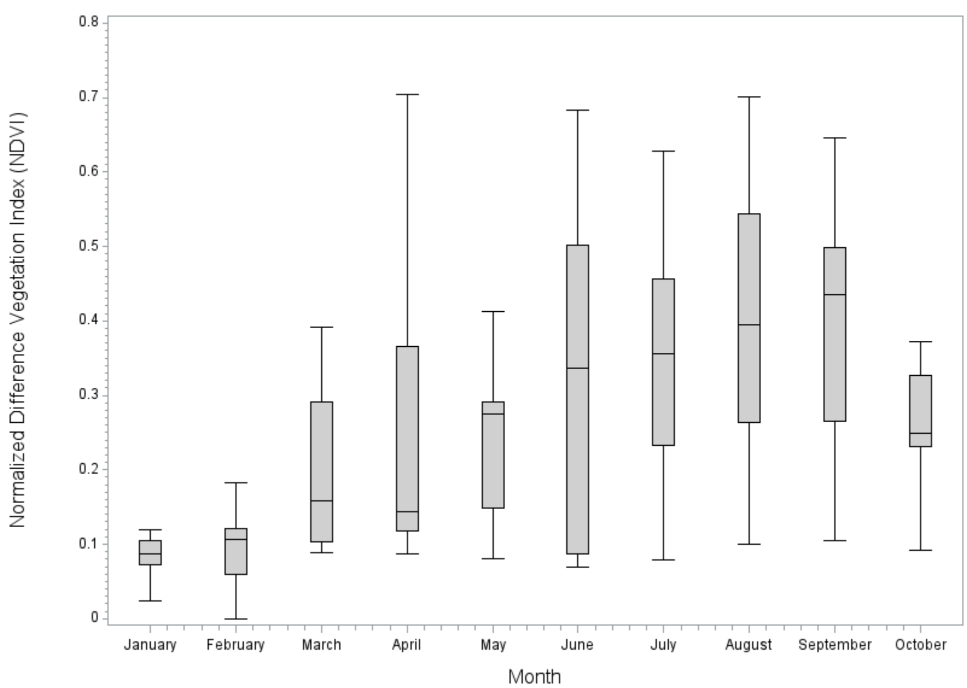


Figure D49 Sudan: NDVI in Kern County, 1990

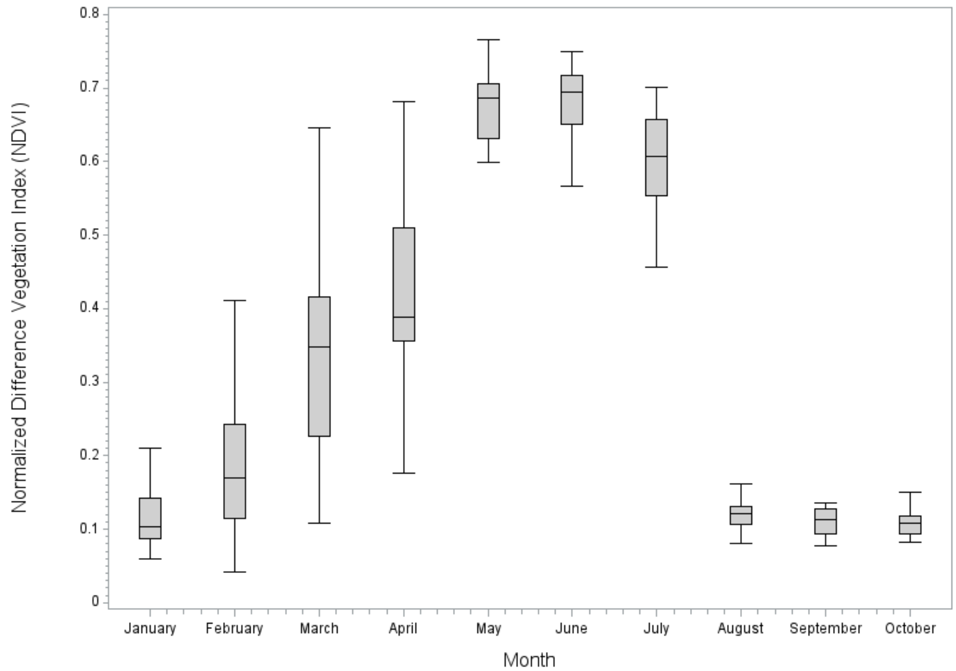


Figure D50 Sugar beet: NDVI in Kern County, 1990

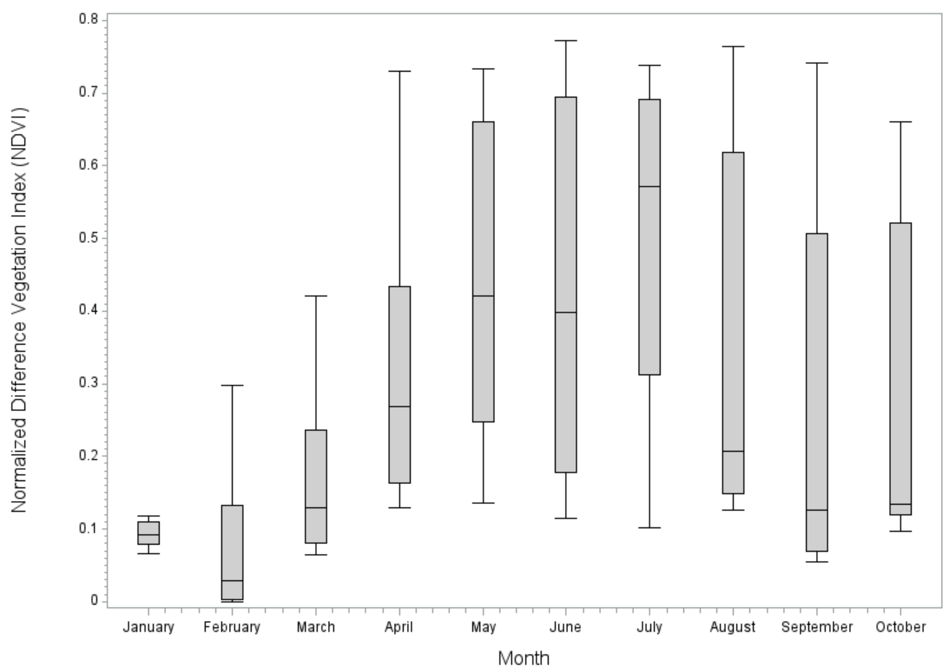


Figure D51 Sweet potato: NDVI in Kern County, 1990

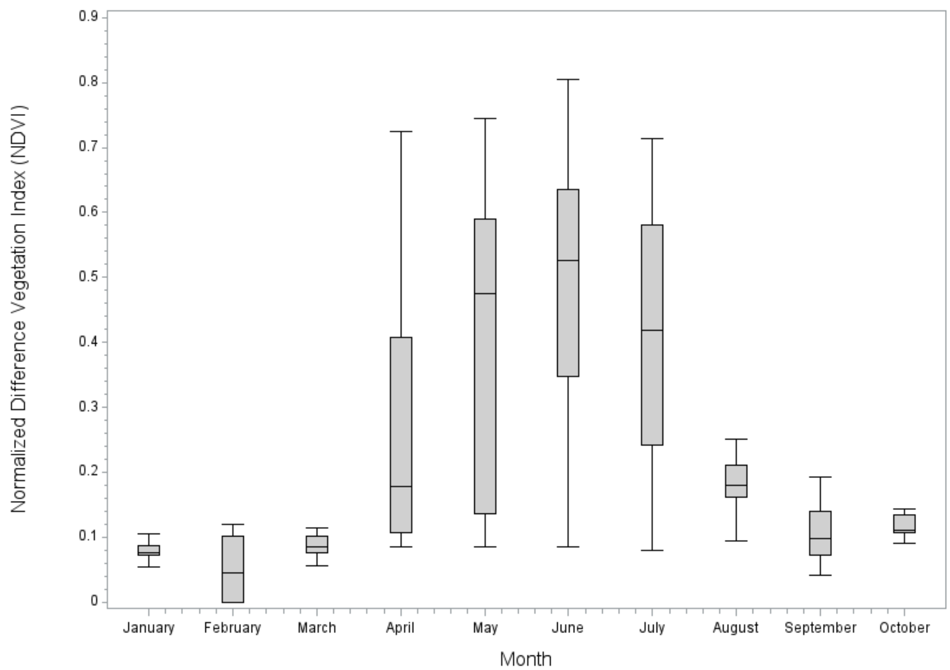


Figure D52 Tomato: NDVI in Kern County, 1990

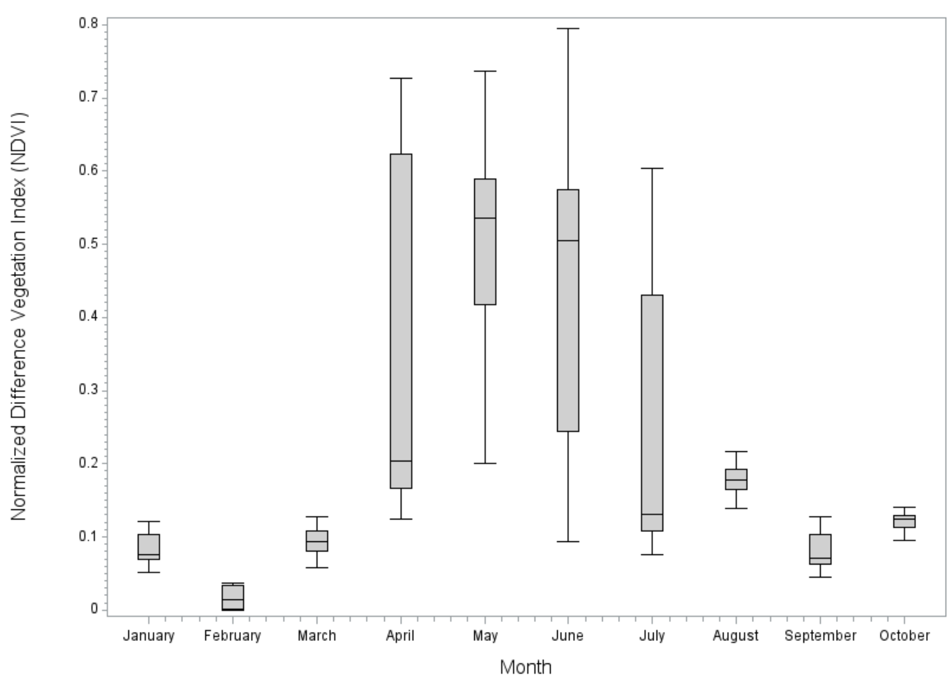


Figure D53 Truck, nursery, and berry crop: NDVI in Kern County, 1990

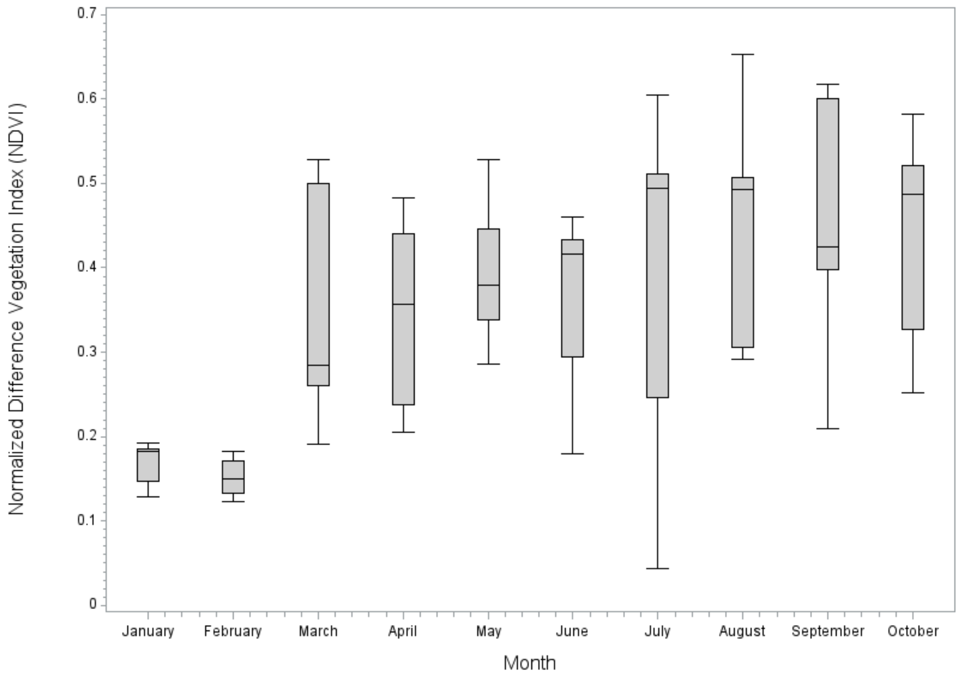


Figure D54 Turf farm: NDVI in Kern County, 1990

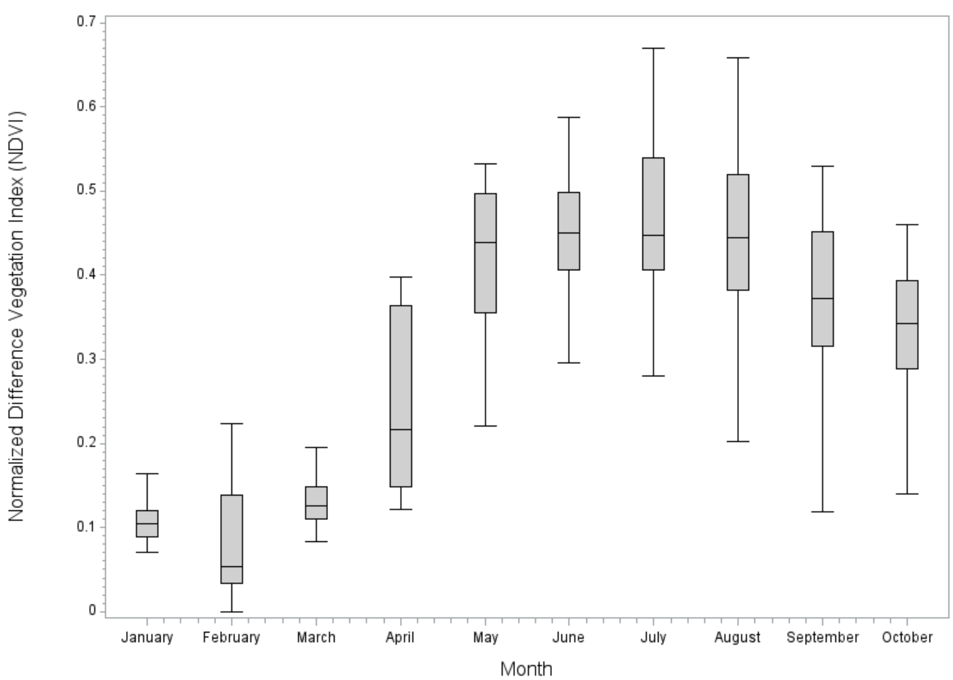


Figure D55 Vineyard: NDVI in Kern County, 1990

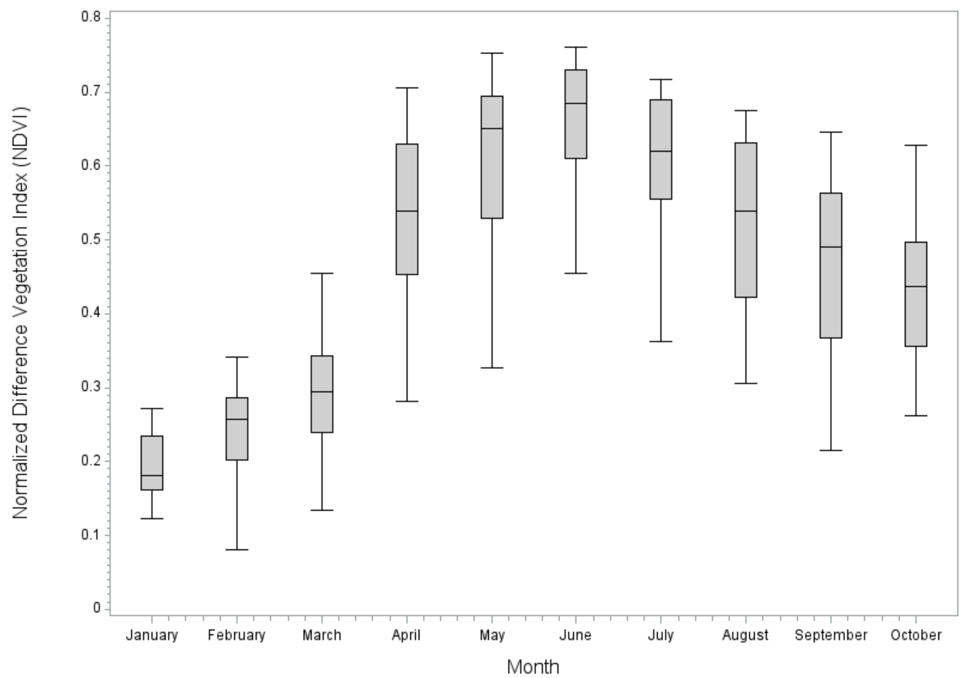


Figure D56 Walnut: NDVI in Kern County, 1990

APPENDIX E: SEGMENTATION AND CLASSIFICATION

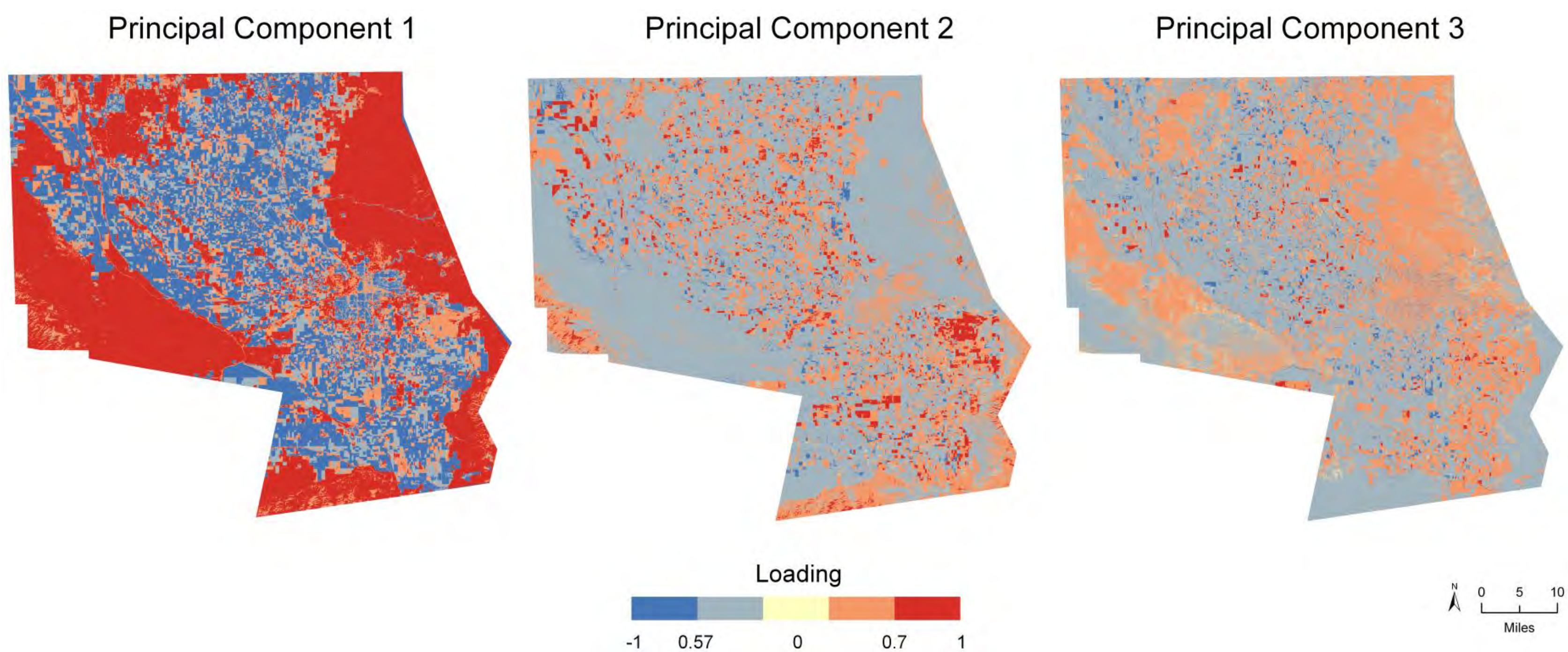


Figure E1 First three principal components derived from 1985 Landsat NDVI images. These PCA images were used for segmentation.

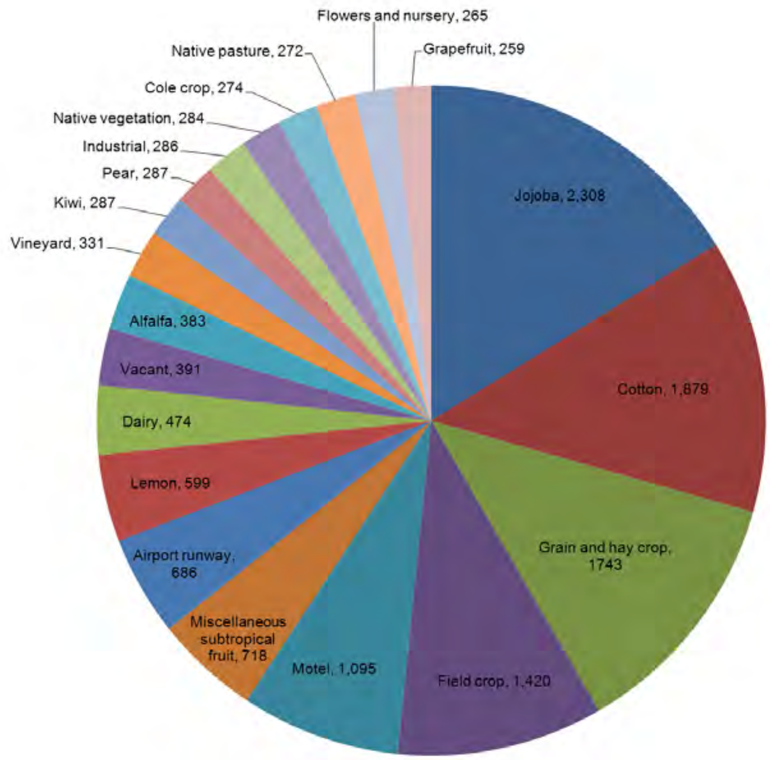


Figure E2 Classification 1 (standard): common land use classes

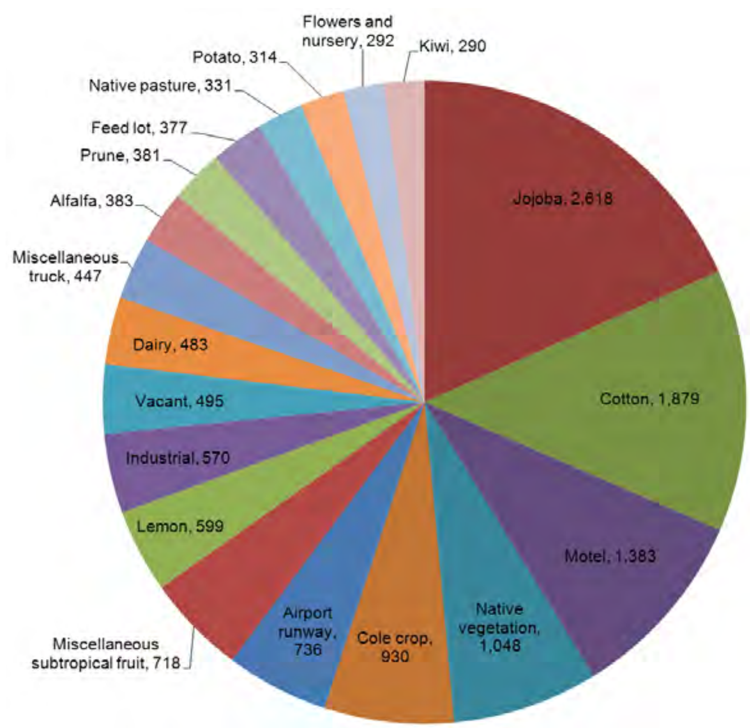


Figure E3 Classification 2 (subclass-required): common land use classes

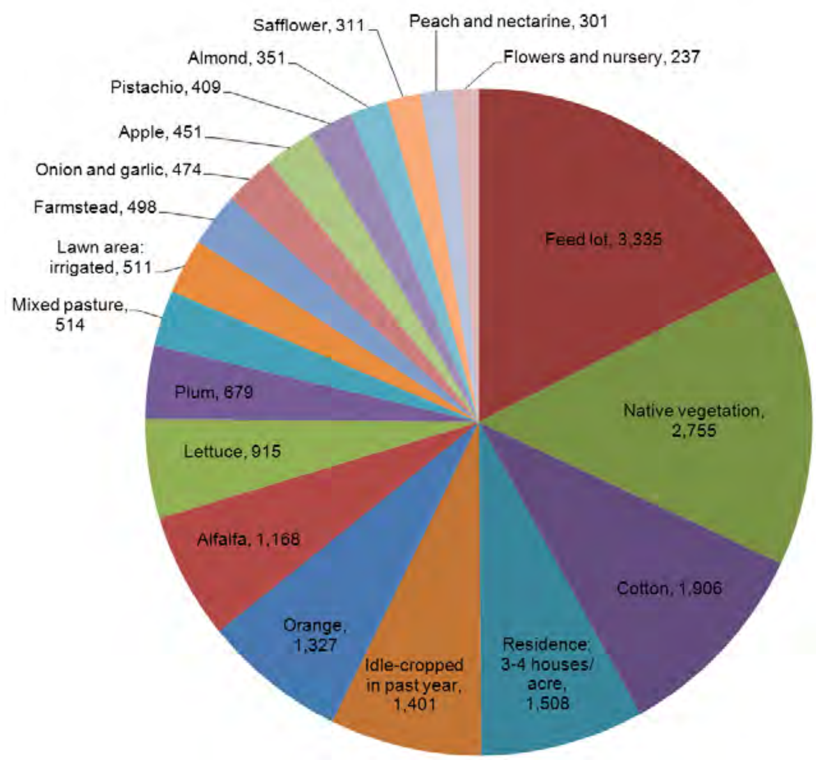


Figure E4 Classification 3 (strict): common land use classes

Table E1 Agricultural land use classes: 1990 Kern County land use survey[1]

Land use class	*N*	Percent	Land use class	*N*	Percent
Cotton	3,036	27.5	Truck, nursery, and berry crop	24	0.22
Alfalfa	1,420	12.86	Turf farm	23	0.21
Field crop	1,226	11.1	Olive	21	0.19
Almond	776	7.03	Pepper	18	0.16
Vineyard	751	6.8	Miscellaneous truck	17	0.15
Grain and hay crop	625	5.66	Sweet potato	16	0.14
Orange	490	4.44	Idle-new lands prepared for crops	13	0.12
Idle-cropped in past year	486	4.4	Grain sorghum	12	0.11
Potato	224	2.03	Jojoba	9	0.08
Onion and garlic	178	1.61	Fig	8	0.07
Pistachio	139	1.26	Asparagus	7	0.06
Mixed pasture	135	1.22	Grapefruit	6	0.05

Table E1 continued

Land use class	***N***	**Percent**	**Land use class**	***N***	**Percent**
Sugar beet	132	1.2	Cole crop	5	0.05
Carrot	130	1.18	Native pasture	5	0.05
Corn	130	1.18	Avocado	4	0.04
Flowers and nursery	128	1.16	Prune	4	0.04
Peach and nectarine	118	1.07	Miscellaneous subtropical fruit	3	0.03
Apple	104	0.94	Bushberry	2	0.02
Bean (dry)	88	0.8	Cabbage	2	0.02
Melon, squash, cucumber	86	0.78	Cherry	2	0.02
Tomato	72	0.65	Strawberry	2	0.02
Plum	60	0.54	Artichoke	1	0.01
Kiwi	45	0.41	Castor bean	1	0.01
Miscellaneous deciduous	39	0.35	Celery	1	0.01
Lettuce	37	0.34	Deciduous fruit and nut	1	0.01
Walnut	37	0.34	Idle	1	0.01
Lemon	32	0.29	Pasture	1	0.01
Safflower	32	0.29	Pea	1	0.01
Bean (green)	25	0.23	Pear	1	0.01
Apricot	24	0.22	Rice	1	0.01
Sudan	24	0.22			

[1] Data from CDWR (2013)

Table E2 Final CSL-classified land use classes

Land use class	***N***	**Percent**	**Land use class**	***N***	**Percent**
Cotton	1,878	18.76	Mixed pasture	116	1.16
Miscellaneous subtropical fruit	717	7.16	Peach and nectarine	100	1.00
Lemon	599	5.99	Apple	99	0.99
Cole crop	525	5.25	Bean (green)	93	0.93
Miscellaneous truck	387	3.87	Orange	80	0.80
Alfalfa	383	3.83	Turf farm	76	0.76
Prune	381	3.81	Sugar beet	75	0.75

Native pasture	329	3.29	Apricot	73	0.73
Potato	311	3.11	Idle-cropped in past year	67	0.67
Flowers and nursery	292	2.92	Onion and garlic	54	0.54
Kiwi	290	2.90	Asparagus	50	0.50
Pear	287	2.87	Safflower	44	0.44
Grapefruit	259	2.59	Corn	40	0.40
Walnut	257	2.57	Grain sorghum	38	0.38
Avocado	201	2.01	Lettuce	35	0.35
Bean (dry)	193	1.93	Sweet potato	32	0.32
Melon, squash, cucumber	192	1.92	Tomato	31	0.31
Sudan	188	1.88	Miscellaneous deciduous	30	0.30
Plum	183	1.83	Fig	28	0.28
Pea	168	1.68	Pepper	15	0.15
Pistachio	168	1.68	Bushberry	14	0.14
Almond	165	1.65	Rice	10	0.10
Carrot	161	1.61	Cabbage	3	0.03
Olive	148	1.48	Idle-new lands prepared for crops	2	0.02
Cherry	141	1.41			

APPENDIX F: APPLIED PESTICIDES AND RURALITY

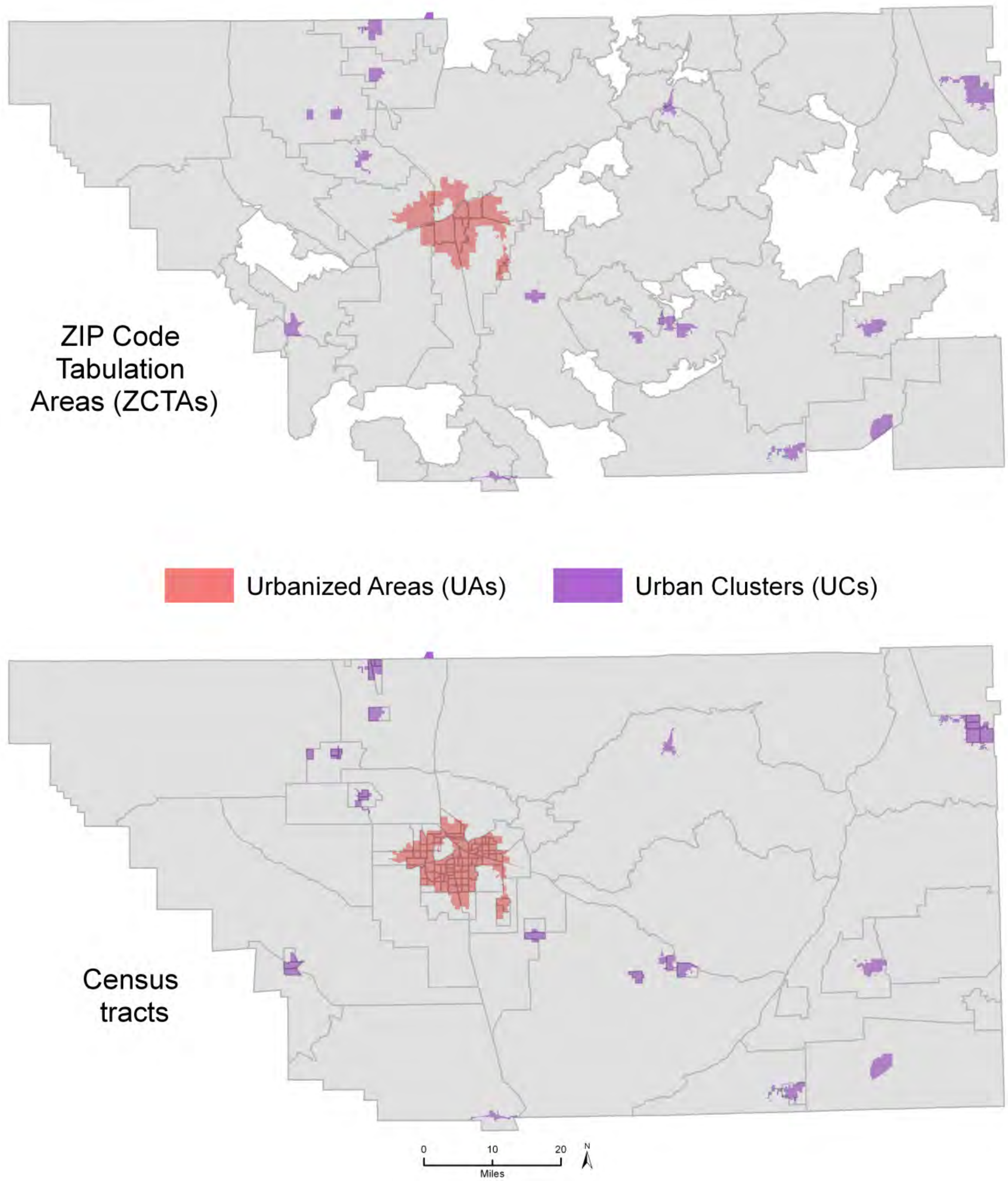


Figure F1 Kern County Urbanized Areas and Urban Clusters (2000)
(Data from U.S. Census Bureau 2013)

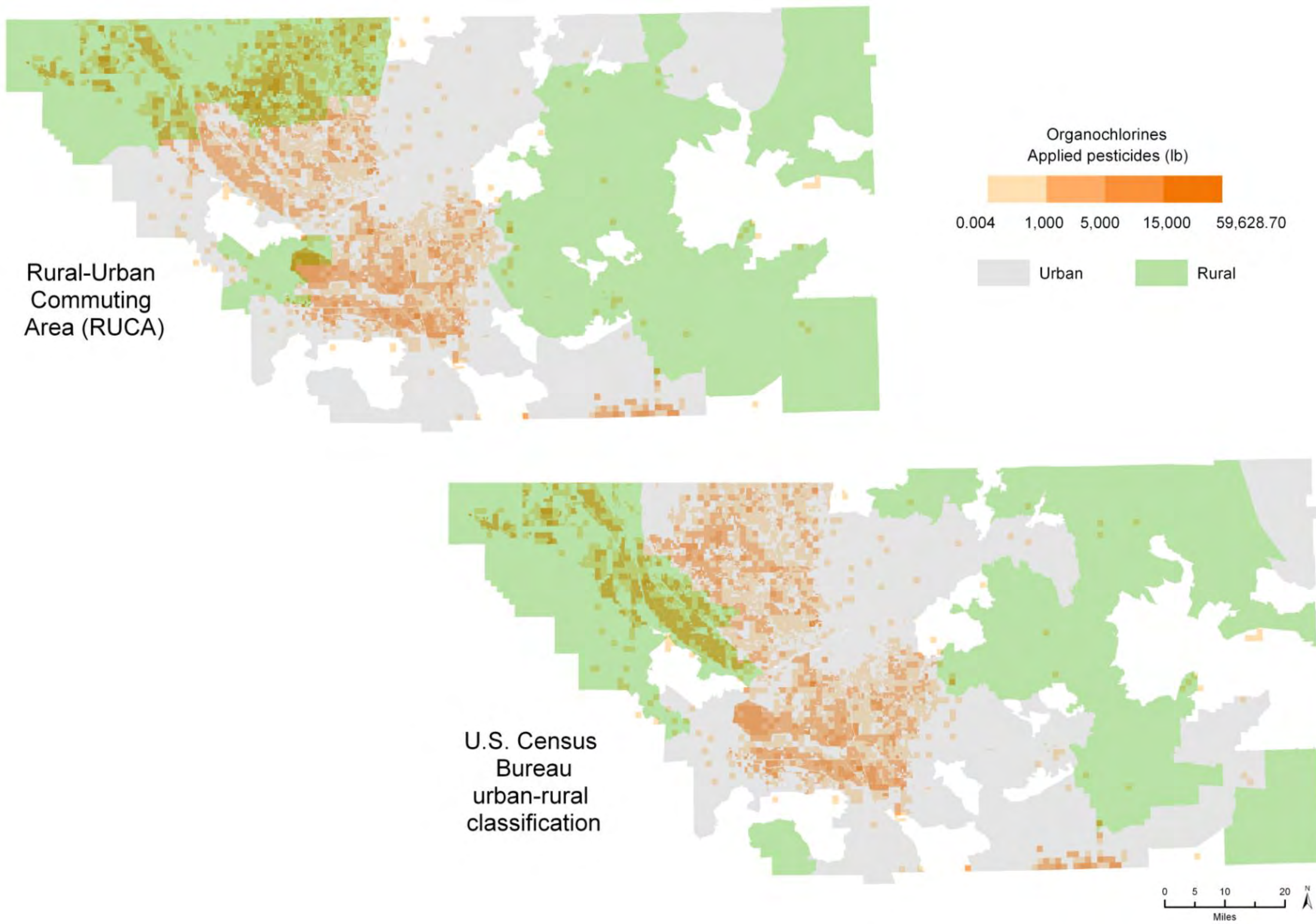


Figure F2 Applied pesticides and ZCTA rurality: organochlorines (Data from CDPR 2013; and U.S. Census Bureau 2013)

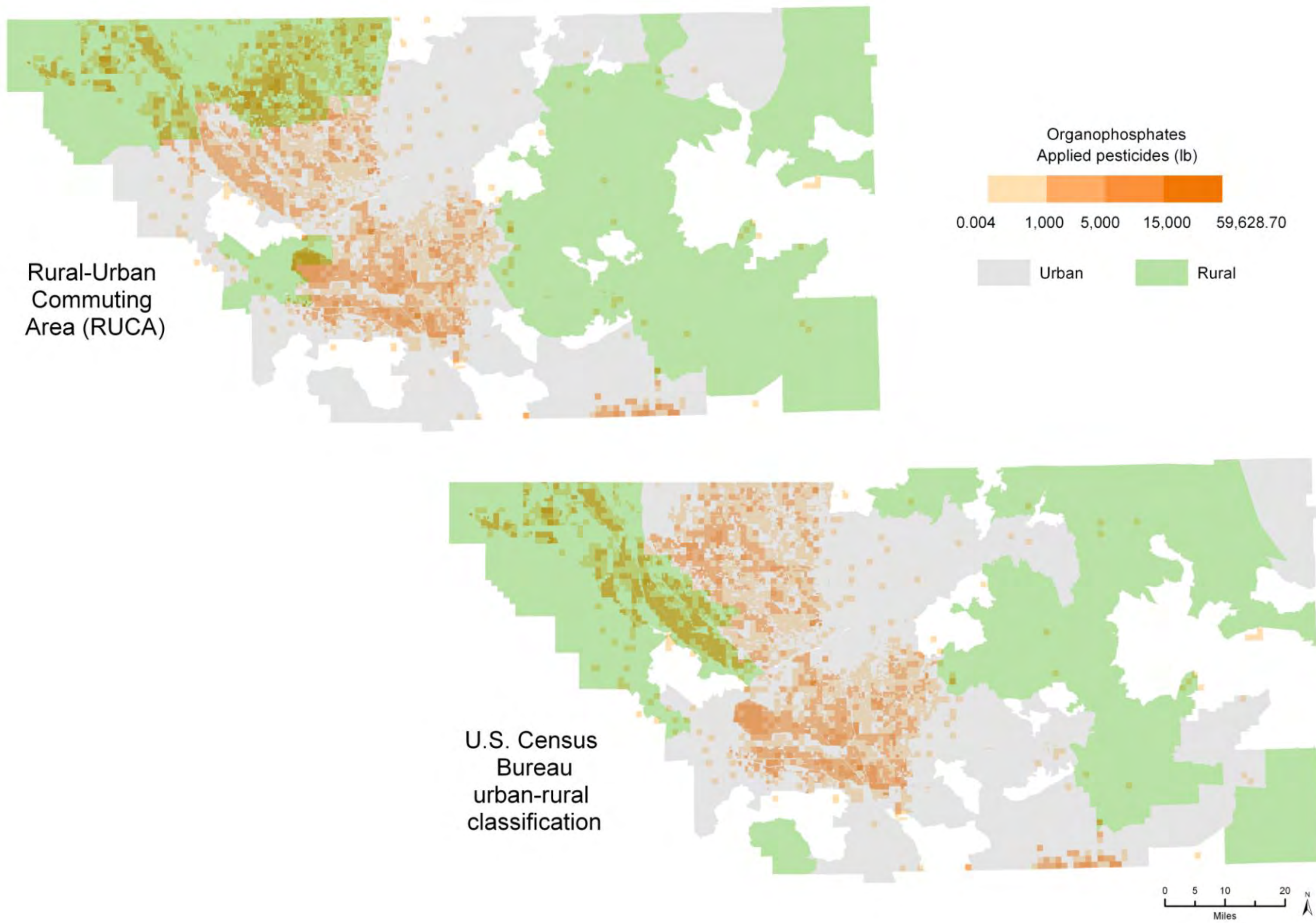


Figure F3 Applied pesticides and ZCTA rurality: organophosphates
(Data from CDPR 2013; and U.S. Census Bureau 2013)

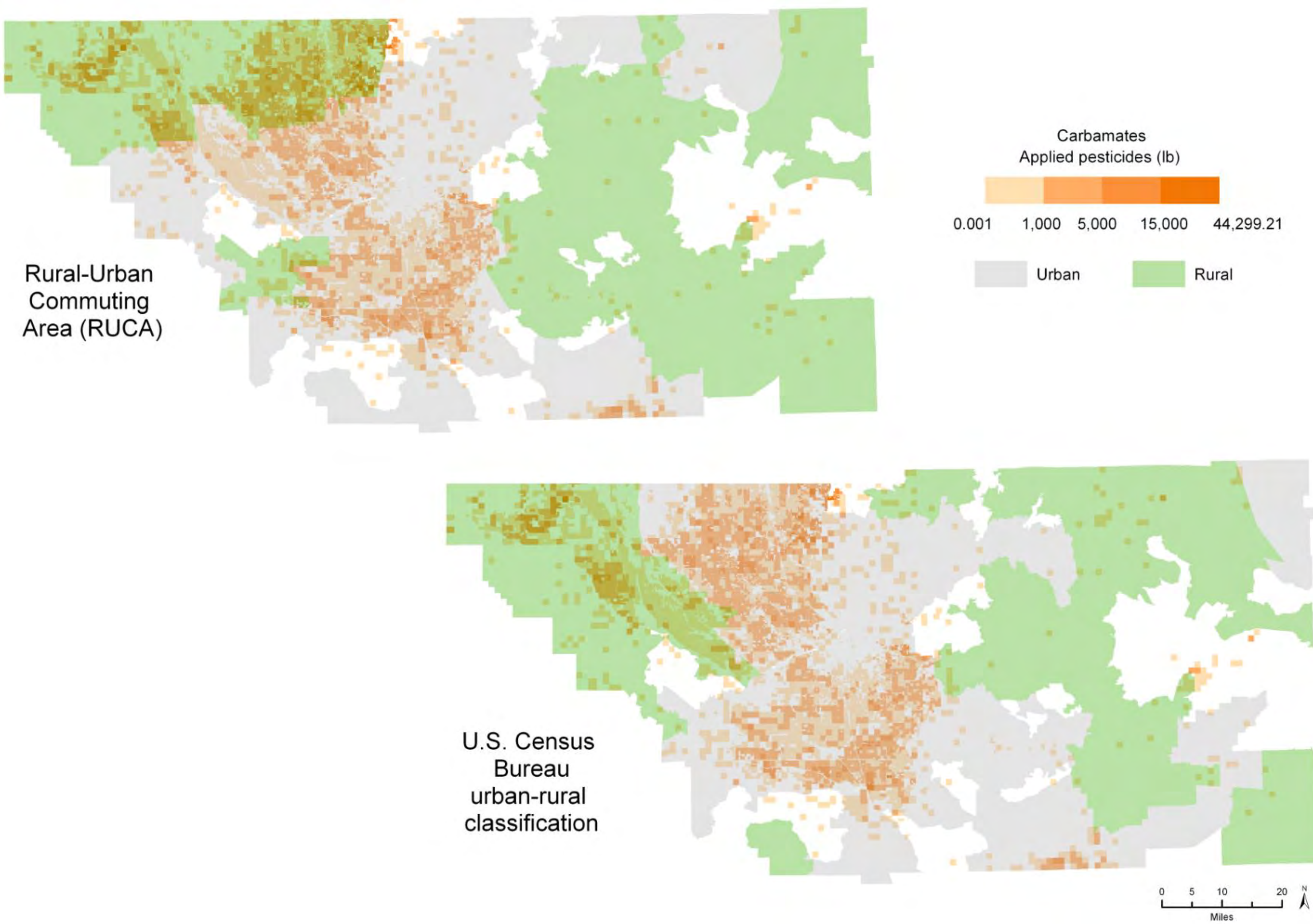


Figure F4 Applied pesticides and ZCTA rurality: carbamates
(Data from CDPR 2013; and U.S. Census Bureau 2013)

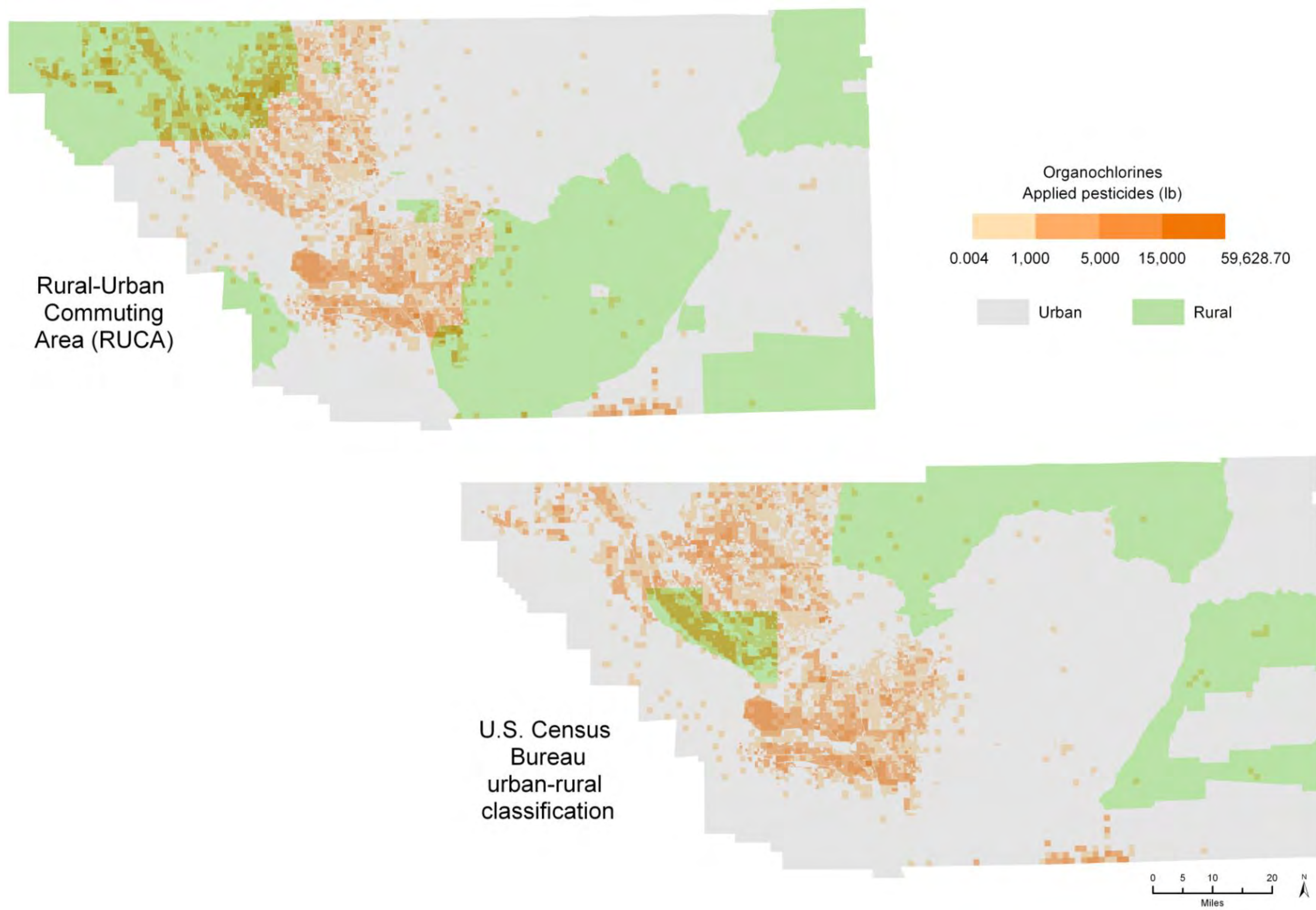


Figure F5 Applied pesticides and census tract rurality: organochlorines (Data from CDPR 2013; and U.S. Census Bureau 2013)

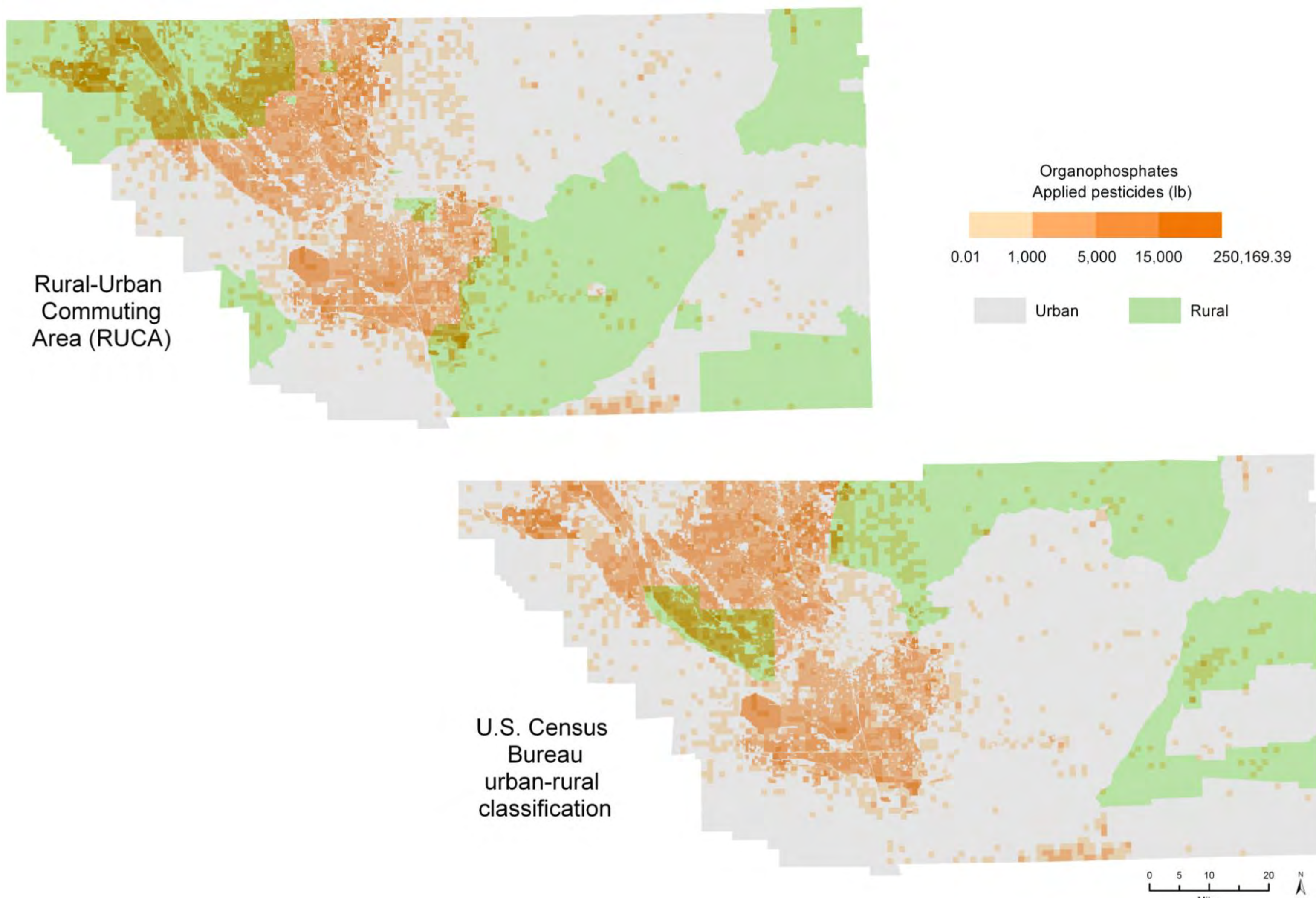


Figure F6 Applied pesticides and census tract rurality: organophosphates (Data from CDPR 2013; and U.S. Census Bureau 2013)

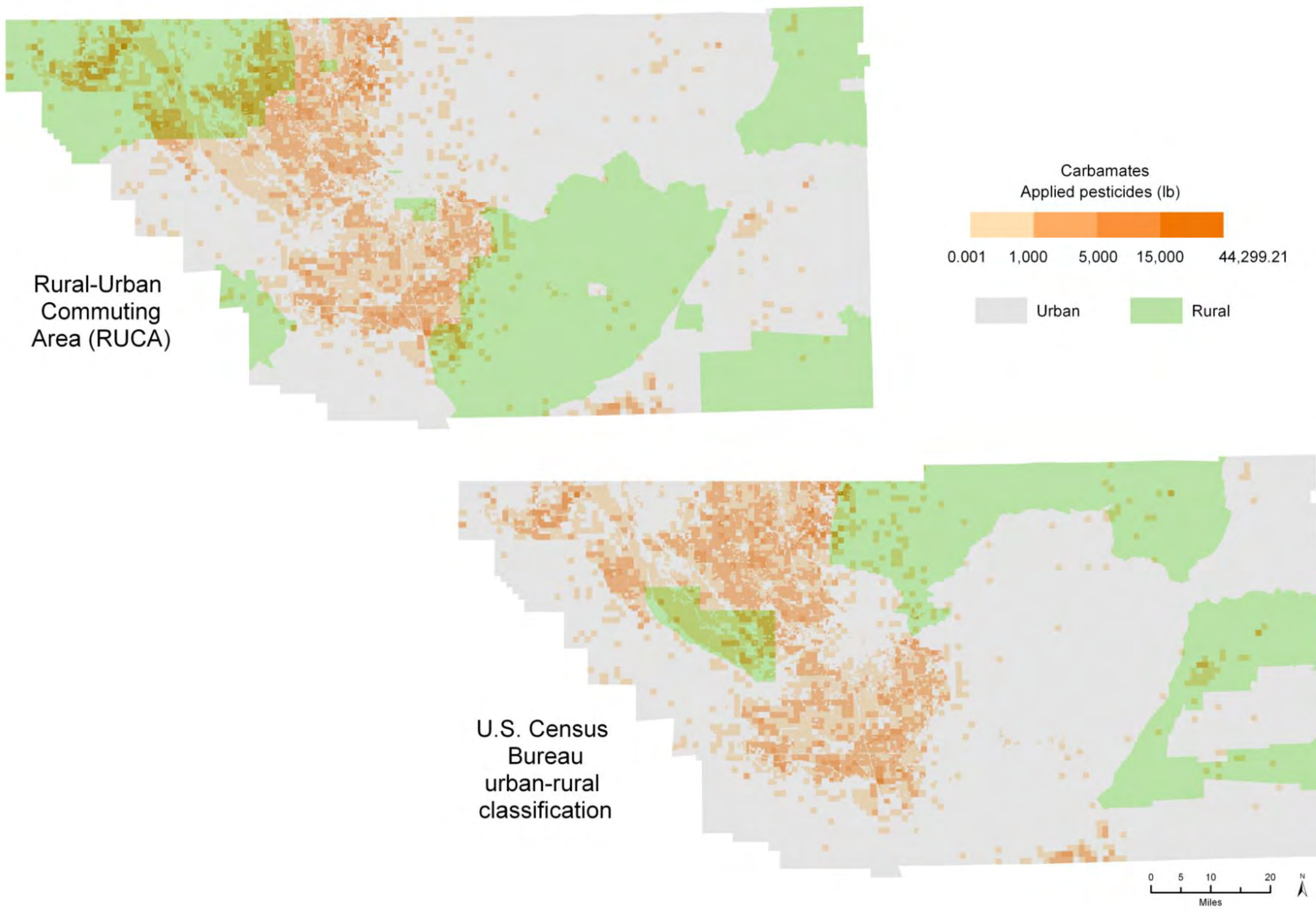


Figure F7 Applied pesticides and census tract rurality: carbamates
(Data from CDPR 2013; and U.S. Census Bureau 2013)

Table F1 Pesticide-treated crop fields and sections intersecting ZCTAs by rurality

Organochlorines						
		N	**Mean ± *SD***	**Median (*IQR*)**	**Min.**	**Max.**
RUCA	Rural	20	54.25 ± 101.69	2 (34)	0	307
	Urban	27	68 ± 110.37	7 (144)	0	457
U.S. Census Bureau	Rural	18	30.06 ± 76.02	1.5 (7)	0	266
	Urban	29	82.07 ± 117.57	14 (155)	0	457
Organophosphates						
RUCA	Rural	20	94.75 ± 163.47	10 (75)	0	481
	Urban	27	104 ± 156.34	21 (206)	0	658
U.S. Census Bureau	Rural	18	54.78 ± 121.67	10.5 (24)	0	481
	Urban	29	128.17 ± 172.45	32 (220)	1	658
Carbamates						
RUCA	Rural	20	76.15 ± 138.84	5.5 (41)	0	415
	Urban	27	87 ± 134.49	12 (187)	0	577
U.S. Census Bureau	Rural	18	42 ± 97.19	9 (15)	0	377
	Urban	29	107.45 ± 149.97	23 (191)	1	577

Table F2 Pesticide-treated crop fields and sections intersecting census tracts by rurality

Organochlorines						
		N	**Mean ± *SD***	**Median (*IQR*)**	**Min.**	**Max.**
RUCA	Rural	28	24.86 ± 93.71	1 (5)	0	486
	Urban	112	21.58 ± 62.88	2 (8)	0	424
U.S. Census Bureau	Rural	7	34.71 ± 72.70	3 (27)	0	198
	Urban	133	21.58 ± 69.81	1 (6)	0	486
Organophosphates						
RUCA	Rural	28	46.93 ± 163.41	4 (6)	0	828
	Urban	112	33.44 ± 89.33	4 (12)	0	631
U.S. Census Bureau	Rural	7	78 ± 101.80	38 (205)	0	239
	Urban	133	33.93 ± 107.86	4 (9)	0	828
Carbamates						
RUCA	Rural	28	35.71 ± 128.7	2 (6)	0	657
	Urban	112	28.29 ± 78.42	3 (11)	0	567
U.S. Census Bureau	Rural	7	57 ± 79.97	33 (118)	0	212
	Urban	133	28.34 ± 90.76	3 (8)	0	657