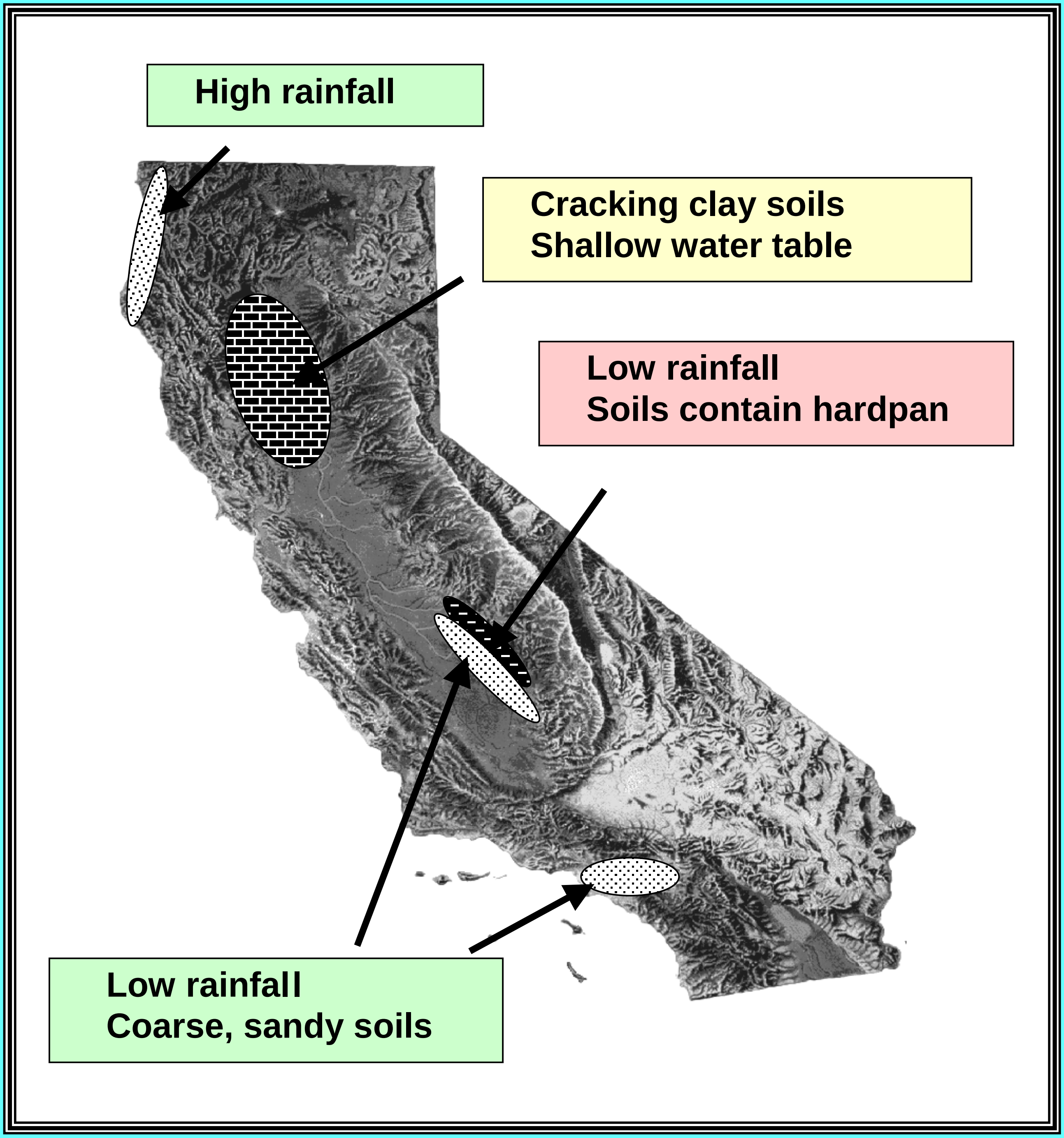
Update of the California Vulnerability Soil Analysis for Movement of Pesticides to Ground Water: October 14, 1999

By

John Troiano, Frank Spurlock, and Joe Marade



STATE OF CALIFORNIA
Environmental Protection Agency
Department of Pesticide Regulation
Environmental Monitoring and Assessment Program
Environmental Hazards Assessment Program
830 K Street, Room 200, Sacramento, Ca 95814 -3510
EH 00-05

Abstract

The CALVUL approach to determining spatial vulnerability to ground water contamination has been previously described (Troiano et al., 1994; Troiano et al. 1997; Troiano et al., 1999). CALVUL is an empirical approach because it attempts to identify similar geographic features amongst sections of land where pesticide residues have been found in ground water. Two unique features of this approach are: 1) that no *a priori* determination is made regarding the pathway for pesticide movement to ground water; and 2) that no relative scale of vulnerability is derived between land areas. This report describes a revision in the clustering analysis of soil data. The revision was conducted because the number of sections with pesticide detections had approximately doubled since the initial development of the CALVUL approach. All of the sections used for this revision originated from DPR investigations which assured that sampled wells had met all aspects of a non-point source determination, especially with respect to visual inspection of well sites. In addition, the soil data tables originally obtained from the National Resource Conservation Service (NRCS) had been updated. The results of this analysis were very similar to the initial clustering analysis. Variables that were important in discriminating clusters were permeability, shrink-swell potential, presence of a hardpan soil layer, and presence of an annual water table. Soil texture in the initial analysis was reflected in values for the No 200 sieve. In this revision, soil texture was indicated by the combination of permeability and shrink-swell potential. Coarse soils were characterized by high permeability values and no shrink-swell potential whereas clayey soils were characterized by very low permeability values and high shrink-swell potential. These observations were compared to the No200 sieve sizes to validate this observation. Although this revision indicated a greater number of clusters, there was better correspondence to general soil maps. The addition of water table as a cluster variable provided greater separation primarily between clayey soils. Presence of a water table could be an important variable in the development of mitigation measures and it is one of the observations that require further investiation.

Acknowledgements

We gratefully acknowledge all the hard work that forms the foundation of this current study, and we thank all the personal in the Environmental Hazards and Assessment Program of the California Department of Pesticide Regulation who have worked on the many studies upon which these data are drawn.

Table of Contents

(Appendix A and B available upon request)

List of Tables

List of Figures

Update of the California Vulnerability Soil Analysis for Movement of Pesticides to Ground Water
October 14, 1999

John Troiano, Frank Spurlock, and Joey Marade

Environmental Monitoring and Pest Management Branch
California Department of Pesticide Regulation
830 K Street, Room 200
Sacramento, CA., 95814-3510
*Corresponding author (jtroiano@cdpr.ca.gov)

Introduction

The reasons for developing a model that identifies areas of land that are vulnerable to ground water contamination are:

1. To increase the efficiency of well monitoring studies. One mandated goal of well studies conducted by the Department of Pesticide Regulation (DPR) is to detect residues for active ingredients that have not yet been detected in California's ground water (Connelly, 1986). Identification of areas with a higher probability of detection should focus monitoring activities and expedite the detection of new residues.
2. To delineate areas where mitigation measures should be implemented. A strength of using a Geographical Information System (GIS) approach is the production of maps that identify areas of land with similar geographic features. If geographic features can be related to higher probabilities of pesticide detection in well water, then mitigation measures could be implemented within delineated areas.
3. To aid in the design and development of mitigation measures. One further step taken in a GIS approach is to relate geographic features to the processes by which pesticide residues move from sites of application to ground water. Practical significance can then be assigned to important geographic features because management practices could be tailored to the delineated areas. Conversely, studies could be designed to determine processes of ground water contamination in areas where further study is needed to describe pathways for movement of residues into ground water.
4. To fulfill programmatic mandates. The U.S. EPA is developing a process for increased regulation of pesticides that have contaminated ground water. States will be required to develop Pesticide Management Plans (PMPs) for pesticides of concern. One prong of the plan is the development of statewide vulnerability assessments. The California Vulnerability approach (CALVUL) is proposed to fulfill this requirement.

The CALVUL approach to determining spatial vulnerability to ground water contamination has been previously described (Troiano et al., 1994; Troiano et al. 1997; Troiano et al., 1999). CALVUL is an empirical approach because it attempts to identify similar geographic features amongst sections of land where pesticide residues have been found in ground water. Two unique features of this approach are: 1) that no *a priori* determination is made regarding the pathway for pesticide movement to ground water; and 2) that no relative scale of vulnerability is derived between land areas. Most other methods focus on delineating land areas where pesticides would leach to groundwater as a result of simple percolation of water from the land surface (National Research Council, 1993). Well sampling studies have been conducted to test spatial indices of vulnerability derived from models based solely on leaching potential (U.S. Environmental Protection Agency, 1992; Balu and Paulsen, 1991; Holden *et al.*, 1992, Kalinski *et al.*, 1994; Roux *et al.*, 1991). Pesticide residues in these studies were detected in wells located in areas identified as relatively invulnerable. This result reinforced our observation that movement of pesticide residues to ground water occurred by multiple pathways, depending on soil, climatic and agronomic factors. In California, pesticide residues had been detected in coarse-textured soil areas on the eastern side of Southern San Joaquin Valley and inland in Southern California (Figure 1.). Leaching with simple percolation is a likely mechanism for pesticide movement to ground water in this area. In contrast, residues have also been detected in areas where leaching was less likely such as in the fine-textured clay soils in the Sacramento Valley. Climatic conditions in areas with detections also vary from relatively dry areas at less than 10 inches annual rainfall to greater than 60 inches of annual rainfall (Figure 1).

In addition to leaching, other potential pathways include: movement of surface water into agricultural drainage wells (Braun and Hawkins, 1991; Roux et al., 1991); movement of water into Karst formations (Hallberg, 1989); or movement of water through cracks in clay soils (Graham et al., 1992). Recently, measurement and modeling the movement of residues in macropore flow has gained more attention (Bergstrom et al., 1991; Chen et al., 1993).

Owing to the uncertainty in determining the exact process of pesticide movement to ground water, we decided to use an empirical approach for the spatial identification of areas that are vulnerable to ground water contamination. The first step in the CALVUL modeling approach was to identify sections of land where pesticide residues had been detected in ground water as a result of non-point source agricultural applications. Data for contaminated wells were obtained from the Well Inventory Data Base that is maintained by DPR (Maes et al., 1992). Sections of land were designated as known contaminated (KC) sections where residues had been detected and attributed to non-point source agricultural applications. Sections are one-square-mile areas of land as described by the USGS Public Land Survey (Davis and Foote, 1966). A section was chosen as the smallest geographical unit because it was also the smallest geographical reference for other databases supported at that time by the DPR such as pesticide use reporting and the Well Inventory Data Base.

A non-point source determination for detection of a pesticide residue in well water is made when 3 conditions are fulfilled.

1. Observation of well construction and pesticide storage and handling around a well rules out potential point sources,
2. Non-point source use of the pesticide in the area near the well is indicated,

3. Residues are measured in more than 1 well within a 1 square-mile area surrounding the original detection.

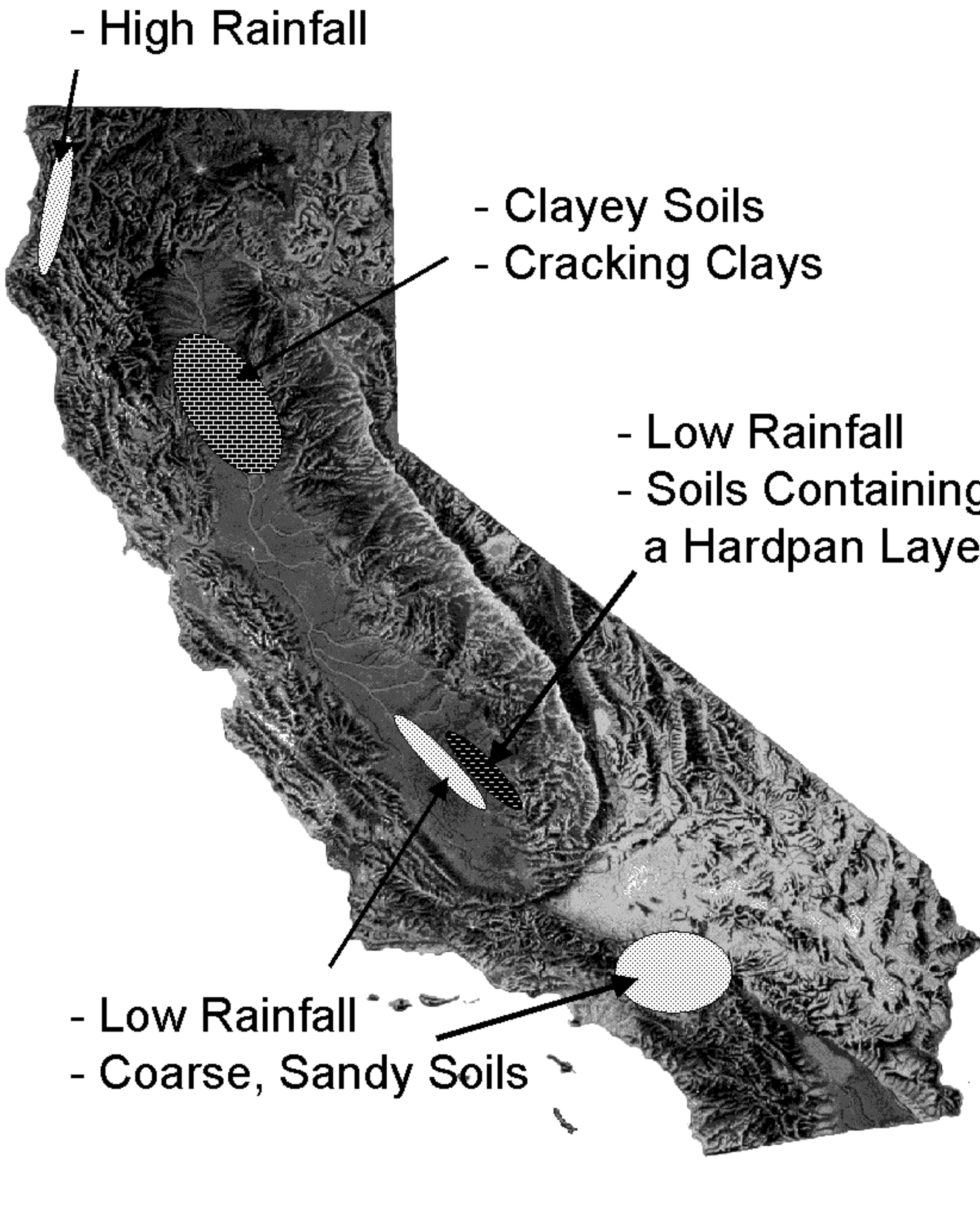
Although most non-point source determinations had been conducted by DPR staff, data reported by the California Department of Health Services (DHS) were from municipal sources, as compared to predominantly domestic, rural wells sampled in DPR studies. Also, many DHS data are for pesticides no longer registered for use in California, i.e. DBCP, 1,2-D, and EDB. Many wells sampled by DHS were determined as non-point source without inspection because municipal wells have a high construction standard, use of many reported pesticides were previously suspended and consequently no longer regulated by DPR, and the integrity of the chemical analysis was high because of regulatory consequences.

Description of Statistical Methodology

Based on the smallest spatial scale as possible, statistical clustering methods were used to identify groups of KC sections using predominant climatic and edaphic features (Troiano et al., 1994). The following discussion focuses on the use of soil variables in the cluster analysis. A forward stepwise cluster selection procedure was developed as based on a suggestion by Fowlkes et al., (1988). In the first step, a cluster analysis was conducted for each separate variable and the single best variable that formed clusters was identified. In the second step, the single best variable was tested in combination with the rest of the variables and the best clustering pair of variables identified. Variables that were highly correlated, at a Pearson Correlation Coefficient $\geq$ 75%, with chosen variables were eliminated from subsequent steps because high correlation between variables tends to inflate statistical measures used to test the performance of the cluster analysis (Aldenderfer and Blashfield, 1984). The stepwise process was repeated until there was no improvement in the statistical criteria used to assess the clustering solution. Statistics used to assess cluster formation were the Cubic Clustering Criterion (CCC), the Pseudo-F and Pseudo-t^2 statistics. Comparison of the absolute value of the CCC both within and between steps was used as the primary indicator for improvement in cluster identification in subsequent steps (SAS Institute Inc., 1983; SAS Institute Inc., 1988).

The geographical significance of the statistical clusters was investigated by plotting the location of the sections that formed each cluster. The clusters did indeed distinguish land areas, providing a statistical reflection of the soil distribution described in Figure 1. In order to apply the results to other geographic areas, a classification algorithm was developed based on the results of the clustering procedure (Troiano et al., 1994). The algorithm was used to classify other sections that lacked pesticide detection or well sampling data into one of the five KC soil clusters or, alternatively, into a not-classified category. At first, the algorithm was based on a Principal Components Analysis (PCA) applied to the results of the clustering analysis. The ability of the classification procedure to determine sections with higher probability of detection was tested in a well sampling study conducted in Fresno and Tulare Counties (Troiano et al., 1997). Wells were sampled in sections of land that had not been previously sampled but that were identified as a member of one of the prevalent vulnerable soil clusters. The rate of pesticide detection in those sections was 43%. This rate was considered successful when

Figure 1. Areas where pesticide residues have been detected in California due to non-point source applications.



compared to results from other surveys that used a similar sampling design of one well per targeted area. This result, however, was dependent on the method used to generate the profiles, whereby a classification algorithm based on Canonical Discriminate Analysis (CDA) appeared more accurate in determining cluster membership than the PCA based algorithm. CDA analysis was suggested by Professor Dallas Johnson (personal communication, Statistics Department, University of Kansas) who is also an instructor for a mulitivariate statistics course given through the Institute for Professional Education, Arlington Virginia. Since the PCA-based algorithm had been developed *ad hoc*, the CDA methodology was chosen because it was a statistically proven method.

Incorporation of Additional Geographic Information

Another feature of the CALVUL approach is that new information can be incorporated into the vulnerability analysis. Statewide data for depth-to-groundwater (DGW) were not available when the project was first initiated, but data were available for Fresno and Tulare Counties. When data from the well study were stratified according to DGW, the probability for detection was approximately 60% in sections with estimated DGW at 50 feet or less, as compared to approximately 10% in sections with DGW estimates deeper than 50 feet Subsequently, a DGW data base has been developed statewide, where available, and it has been included as a geographical data layer to indicate areas with a greater potential for detection of pesticide residue (Troiano et al., 1999).

Testing an Application of the Vulnerability Classifications

The utility of the CALVUL approach was further evaluated in a study designed to determine the presence of norflurazon residues in California's ground water (Troiano et al., 1999). One of the regulatory objectives of DPR's ground water program is to conduct retrospective well sampling studies to determine the presence of new active ingredients in California's ground water. Norflurazon was chosen as a candidate active ingredient because it is a pre-emergence herbicide that is a potential substitute for simazine and bromacil. Simazine and bromacil are both commonly detected in well studies, especially in sampling conducted in Fresno and Tulare Counties. Norflurazon is a pre-emergence herbicide that exhibits physical-chemical properties similar to simazine and bromacil. Norflurazon is persistent in soil with an aerobic soil half-life of approximately 90 days, is not strongly sorbed to soil with a Koc of approximately 600, and is soluble in water at 28 ppm. These values are within the range in values for physical-chemical properties of other known ground water contaminants. The ranges are 8-1,000 days for aerobic soil-half life, 6-7,100 for Koc, and 0.6-780,000 ppb for water solubility (Johnson, 1991).

In the retrospective well study for norflurazon, residues were detected in 8 of 32 wells sampled in Fresno County. This result was significant because residues had not been detected in ground water studies for 18 other pesticides conducted under a protocol that included toxicity in prioritizing candidate pesticides. The success in detection of norflurazon was attributed to; 1) use of the new protocol placing greater emphasis on mobility and persistence of the candidate pesticide, and 2) use of the CALVUL model results to focus sampling in areas with a higher probability of pesticide movement to ground water.

Revision of the Soil Classification and Further Application

The DPR is proposing to implement the CALVUL model through regulations that will increase the preventative aspects of our program. This report describes:

1. An update of the clustering and profiling analyses for the soil variable portion of CALVUL;
2. Comparison of CALVUL results to digitized soil data for Tulare County;
3. The application of the CALVUL model results statewide. The model will be used to determine where permits will be required for agricultural use of pesticides listed as 6800 (a) active ingredients according to the Pesticide Contamination Prevention Act (PCPA) (Connelly, 1996). Pesticide active ingredients listed in section 6800 (a) of the Department of Food and Agriculture code are regulated because they have been detected in ground water due to non-point source agricultural use.

The revision in the clustering analysis has been conducted because the number of sections with pesticide detection has increased in the Well Inventory Dat a Base and the soil data tables originally obtained from the National Resource Conservation Service (NRCS) have been updated. Also, all of the sections used for this analysis originate from DPR investigations which assures that sampled wells have met all aspects of a non-point source determination, especially with respect to visual inspection of well sites.

Materials and Methods

Determination of Known Contaminated Sections

Data for pesticide detections in well water were obtained from the Well Inventory Data Base maintained since 1985 by the California Department of Pesticide Regulation (DPR) (Maes *et al.*, 1992). The Pesticide Prevention Contamination Act (Connelly, 1986) requires the DPR to determine whether or not reported detections are due to legal agricultural use. Subsequently, a Known Contaminate (KC) section was defined as a section where pesticide residues had been found in well water due to legal agricultural use (Appendix A, SQL program #1 page A-1). The pesticide active ingredients and breakdown products detected in well water in KC sections are listed in Table 1. DBCP detections, though numerous, were omitted from the study. Use of DBCP was banned in 1979. Since then, a large number of detections in well water have been reported, primarily from sampling conducted by the California Department of Health Services (Brown *et al.*, 1986). Detection could have resulted from movement of contaminated ground water between sections during the time span between cessation of use and sampling of well water. Although this problem may exist with other detected pesticides, DBCP represented an extreme in terms of spreading in ground water aquifers due to the widespread high rates of application, high mobility of volatile fumigants such as DBCP, and the extraordinarily long half-life of DBCP, which is estimated at greater than 100 years (Burlinson *et al.*, 1982). The less extensive, lower rates of application and shorter half-lives for the other pesticide ground water contaminants provide some assurance that detection of these pesticides are more reflective of local use.

Table 1. Active ingredients detected in California well sampling investigations and determined to be derived from non-point source applications.

Common Name	Action	Number of Sections with Detection
Atrazine	Pre-emergence Herbicide	59
Bentazon	Pre-emergence Herbicide	48
Bromacil	Pre-emergence Herbicide	132
Deethylatrazine	Atrazine Breakdown Product	16
Deisopropylatrazine/deethylsimazine	Atrazine and Simazine Common Breakdown Product	84
Didealkylated Triazine	Atrazine and Simazine Common Breakdown Product	5
Diuron	Pre-emergence Herbicide	220
Norfluazon	Pre-emergence Herbicide	3
Prometon	Pre-emergence Herbicide	20
Simazine	Pre-emergence Herbicide	314
2,3,5,6-tetrachloroterephthalic Acid	Breakdown Product of Chlorthal-Dimethyl which is a Pre-emergence Herbicide	15

Soil Data
Data for physical and chemical properties of soil were obtained at the level of soil mapping unit as delineated in the county soil surveys of the Natural Resources Conservation Service (formerly the USDA Soil Conservation Service) (Soil Conservation Service, 1983; Soil Survey Staff, 1997). The type of mapping unit used in this study was primarily surface texture phases of consociations of soil series. Since digital data for soil mapping units were not available, a data base was developed that catalogued all soil mapping units by Township/Range/Section (T/R/S) for all published soil surveys. This data set was named the California State Mapping Unit Identification (CSMUID) and it was developed through contract with Tom Rice at Cal Poly, San Luis Obispo initiated by Bob Teso (personal communication, formerly with DPR at the University of Riverside, Riverside, CA). The data set was augmented with preliminary data from soil surveys not yet published such as Glenn County, the Western Part of Tulare County, and Kern County. Soil mapping units in the augmented CSMUID data set were matched to the KC sections extracted from the Well Inventory Data Base. After soil mapping units were assigned to KC sections, data for the physical and chemical properties for each mapping unit were extracted from the National Map Unit Interpretations Record (MUIR) Database provided by the USDA-NRCS Soil Survey Division. The matching of CSMUID data to KC sections and assignment of MUIR data was executed in a single SQL program (Appendix A, SQL program #2 page A-2). The MUIR data was contained in two databases, one named the COMP for composition and the other named LAYER. These databases contain estimates for chemical, textural, and observational data. Data in the COMP table are related to the entire soil column whereas data in the LAYER table are related by soil layer down to the 1.5 meter depth for each soil mapping unit. Both are available through the Internet at http://www.statlab.iastate.edu/cgi-bin/dmuir.cgi and they were downloaded on April 22, 1999.

One other step that was included in this revised analysis was to weight the values according to the percent composition of MUID soil components. This suggestion was made during discussion of the previous results with Dr. Minghua Zhang, UC Davis. Some soils are a complex with a percentage of each specified in the COMP table. For example, the Cometa-San Joaquin sandy loam MUID (CzcB) is composed of 60% Cometa series and 35% San Joaquin series with the remaining 5% a mixture of non-defined soils. Data for this MUID was weighted according to the percentages indicated for the Cometa and San Joaquin soil series.

Statistical Analysis

Data Preparation
Many of the variables are descriptive in nature such as classification for shrink-swell potential as low, moderate, or high. These ordinal variables in the MUIR database were transformed to a numeric scale (Appendix A, SAS program #3a page A-7 and program 3b page A-10). For numeric variables, high and low values were reported for each variable so mid-points were calculated for this study. In the initial analysis reported in 1994 (Troiano et al., 1994), descriptive variables for soil texture in the MUIR database

were transformed to a numeric scale by assigning values for sand and clay from the centroid of corresponding textural classes in the Soil Triangle (Soil Conservation Service, 1975). Owing to the high correlation between the derived texture variables and data for sieve sizes, derived texture data from descriptive variables were considered highly redundant and they were not included in this analysis (see Table V in Troiano et al., 1994).

Variables were derived to partition the soil layer data between surface and subsurface conditions. For the surface soil layer, variables were calculated to represent a sectional value by averaging data from the first soil layer over all soil mapping units within a section. The depth of the soil layer is dependent on the soil horizon and is not consistent for all soil MUIDs. For the subsurface soil layer, variables representative of a section were derived by averaging data for all soil layers below the first layer within a mapping unit with each value weighted according to the depth of each layer. The weighted averages were then averaged across all mapping units within the section. The number of sections with sufficient data for use in the statistical analysis was 465. Although there were 519 KC sections, contemporary soil survey data was lacking for some KC sections in Del Norte, Humbolt, Los Angeles, Orange, and San Bernardino Counties.

One other rule used in the analysis was the exclusion of mapping units with slope greater than or equal to 15% for the high value (Troiano et al., 1994). The exclusion of these data was based on a previous observation that slopes with these values did not represent agricultural areas where contamination had occurred. Within a section, agricultural cropping patterns that could contribute to ground water contamination abutted these sloped areas. Averaging the data for these two conditions would produce estimates that were not representative of the soil used for agricultural purposes.

As a measure of the accuracy of the model estimates, a comparison was made between the CALVUL estimates and digitized soil data for Central Tulare County (Soil Survey #660). This level of data is denoted as the SSURGO set of data, which indicates that the base maps for individual soil surveys have been digitized. Only a few have been digitized. The data were downloaded from the internet address at http://www.ftw.nrcs.usda.gov/stssaid.html and processed using ArcView 3.1 Geographical Information Systems software Environmental Systems Research System (ESRI, Redlands, CA). Two themes were created where a Public Land Survey Section (T/R/S) theme was presented over the NRCS Soil Series CA660 (Central Tulare County) theme. Both themes were in Albers projection units. The Tabulate Area Function in the ESRI Spatial Analyst extension version 1.1 software was used to perform a cross tabulation of the area between the NRCS digitized soil layer and the T/R/S section layer theme. These areas were exported in DBF 4.0 format and then imported into Microsoft Excel 5.0. A summary table was derived that contained the percentage of each soil MUID found within each T/R/S in the Central Tulare County soil survey.

Cluster Analysis

The forward stepwise cluster method has been previously described (Troiano et al., 1994). It was based on a forward selection technique suggested by Fowlkes et al. (1988). Prior to analysis, variables were standardized to mean 0 and standard deviation 1 to remove effects of scale. In the first step, the single best clustering variable was identified using statistical criteria. In the second step, the single best variable was tested in combination with the rest of the variables and the best clustering pair of variables identified. As previously described, variables that were highly correlated with chosen variables were not included in subsequent steps because correlation between variables tends to inflate statistical measures used to test the performance of the cluster analysis (Aldenderfer and Blashfield, 1984). The number of KC sections was large enough to produce statistical significance at low values of the correlation coefficient, so variables with Pearson correlation coefficients ≥ 0.75 were considered highly correlated (Appendix A, SAS program #4 page A-13). The stepwise process was repeated until there was no or only marginal improvement in the statistical criteria used to assess the clustering solution (Appendix A, SAS programs #5 page A-14). The data used for the analysis is reported in Appendix A page A-45.

Three statistical measures were used to determine the number of clusters; the Cubic Clustering Criterion (CCC), the Pseudo-F and Pseudo-t^2 statistics (SAS Institute Inc., 1983; SAS Institute Inc., 1988). As suggested in the SAS publication, CCC values above 3.0 were considered indicative of cluster formation in step 1. An increase in the CCC value was used as the indication for cluster building in subsequent steps. Peaks in the Psuedo-F and valleys in the Psuedo-t^2 statistics are indicative of cluster formation. Two clustering methods were used: Average linkage and Centroid. In the Average method, distance between two clusters is computed as the average distance between pairs of observations, one in each cluster. In the Centroid method, distance between two clusters is computed as the squared Euclidean distance between their centroids. The appropriate number of clusters at each step was determined as the best level of agreement between statistical criteria and between methods.

Profiling Algorithm

The classification method was based on Canonical Discriminant Analysis (CDA) and it was described in Troiano et al., (1997). A CDA analysis was conducted on the data output from the final clustering step. This data set contained the T/R/S identifier, the cluster identification number assigned to each section, and the corresponding raw data for each of the variables identified in the final step. In the CDA analysis, the variable identifying the cluster was designated as the class variable and the soil variables were designated as the explanatory variables (Appendix A, SAS program #6a page A -62). The CDA produces Canonical Variates (CVs) that identify the location of the clusters in canonical space (SAS Institute Inc., 1988).

The classification algorithm was defined as the CV coefficients produced from the CDA that can be applied to raw data (Appendix A, SAS program #6b page A-63, SAS program #6c page A-64, and SAS program #6d page A-66). A section was classified as a member or non-member of a soil cluster by calculating the Euclidean distance between the canonical variate coordinates developed for each section and the centroid mean of each KC soil cluster. This value was compared to the Euclidean distance calculated for the radius of the 95% population tolerance interval constructed around the centroid of each cluster (personal communication, Professor Dallas Johnson, Department of Statistics, Kansas State University). For 2 CVs, the interval was defined as a circular population tolerance interval that was constructed around the centroid means for each of the KC soil clusters. The radius was determined as $\sqrt{}(\chi^2_{0.05}/n)$ with χ^2 = the number of CVs and n=1. For 2 CVs the value was $\sqrt{}(5.99/1)=2.447$. In the previous report (Troiano et al., 1997), 2 CVs were adequate. In this revision, 3 CVs were also tested. For 3 CVs, spherical population tolerance intervals were constructed around each cluster's centroid with the radius determined as $\sqrt{}(7.81/1)=2.795$. In the case of multiple cluster membership, the section was considered a member of the cluster with the smallest Euclidean distance to the cluster centroid. This procedure retained the possibility of producing not-classified sections because the coordinates for a candidate section could fall outside the canonical space defined all of the cluster centroids and their tolerance intervals. An error rate was determined by comparing cluster assignments between a reclassification of the original KC data according to the profile algorithm and the cluster analysis.

Results and Discussion

Stepwise Clustering of Soil Variables

The acronyms and brief description of soil variables used in the statistical analyses are reported in Table 2 (Soil Survey Staff, 1997). Results for the correlation analysis were very similar to the previous study (Table 3 in this report compared to Table 2 in Troiano et al., 1994). There was a group of variables derived from the LAYER table that were highly correlated ($R^2 \geq 0.75$) and where soil texture measurements had been conducted on soil material less than 75 mm in diameter. These variables were sieve sizes 40 and 200, clay content, shrink-swell potential, permeability, bulk density, and cation exchange capacity. In most cases there was a high correlation between the values derived for the surface and subsurface variables. It is interesting that available water holding capacity of the surface layer (AWC1) appeared correlated with some of the texture variables whereas the subsurface values (AWC2) were relatively uncorrelated with all other variables. Variables for coarser sieve sizes, No 4 and 10, were not correlated with the texture variables but they were related to one another. The remaining variables, except for hydrologic group, were not highly correlated. These variables had been primarily derived from the Composition (COMP) table with the exception of the indicators for rock fragments, organic matter content, and soil salinity. Lastly, data for hydrologic group (HYD) was highly correlated with some of the texture variables.

The distribution of a few of the variables indicated that they might not be useful in the cluster analysis (Appendix B1, Univariate Analysis page B-1). For example, the distribution of percent rock fragments in soil was very skewed, as indicated by the Inch101 variable, which had only 12 observations greater than zero and with the positive values relatively evenly distributed from each other. Variables with problematic distributions were Inch31, Inch32, Inch101, Inch102, and Bedrock.

Table 2. Description of soil variables used in cluster and profiling analyses with subfixes 1 and 2 referring to surface and subsurface soil layers, respectively.

Variable Acronym	Description of Variable
Derived from Layer Table	
Awc1, Awc2	Available water holding capacity of soil
BD1, BD2	Bulk density of soil less than 2mm in diameter
Clay1, Clay2	Measured clay content in soil
CEC1, CEC2	Cation exchange capacity of soil
Inch101, Inch102	Weight percent of whole soil greater than 250 mm (10 inches) .
Inch31, Inch32	Weight percent of whole soil greater than 75 mm (3 inches) and less than 250 mm (10 inches)
No41, No42	Percent by weight of soil material smaller than 75 mm in diameter that passes a no. 4 (75 mm) sieve
No101, No102	Percent by weight of soil material smaller than 75 mm in diameter that passes a no. 10 (2 mm) sieve
No401, No402	Percent by weight of soil material smaller than 75 mm in diameter that passes a no. 40 (.425 mm) sieve
No2001, No2002	Percent by weight of soil material smaller than 75 mm in diameter that passes a no. 200 (0.074 mm) sieve
Om1, Om2	Percentage of organic matter in soil
Perm1, Perm2	Permeability measurements
Salin1, Salin2	Measure of soil salinity
Shrink1, Shrink2	Shrink-swell potential of soil with ‘low’, ‘moderate’, and ‘high’ categories coded as 0,1, and 2, respectively
Derived from Composition Table	
Bedrock	Indicator for bedrock material within 1.5 m depth with ‘HARD' or 'SOFT' =1 otherwise 0
Drain	Drainage class identifies the natural drainage condition of soil coded with 'VP'=0;‘P'=1;'SP'=2;'MW'=3;'W,MW'=3.5;'W'=4;'W,SE'=4.5;’SE'=5;'SE,E'=5.5;'E'=6;'P,E'=3.5;
Flood	Flooding indicates the temporary covering of the soil surface by flowing water with ‘NONE' and 'RARE'=0 and 'COMM', ’FREQ', and 'OCCAS'=1
Hyd	Hydrologic group identifying similar runoff potential under similar storm and cover conditions with groups ‘A’ and ‘B’=0 and groups ‘C’, ‘D’, and ‘C/D’=1
Pan	Indicator for cemented hardpan with none=0 and soil with a hardpan=1
Slope	Surface slope of soil
Wattab	Indicator for presence of a water table above 1.5 m and if 'APPAR' or 'PERCH' then wattab=1 otherwise wattab=0
Watsoil	Indicator for hydric soil condition and if hydric='Y' then watsoil=1 otherwise watsoil=0

Table 3. Correlation matrix for soil variables. At n=465, a Pearson correlation coeffiecient of 0.13 is significant at p=0.01 so coefficients of 0.75 or greater are underlined to illustrate trends in data. Acronyms are defined in Table I.

Pearson Correlation Coefficient (n=465

	CORRELATED WITH SOIL TEXTURE																CORRELATED WITH SOIL COARSENESS				NOT HIGHLY CORRELATED															
	Clay1	No2001	No401	Shrink1	Perm1	BD1	CEC1	Clay2	No2002	No402	Shrink2	Perm2	BD2	CEC2	AWC1	AWC2	No41	No101	No42	No102	Inch101	Inch31	Inch102	Inch32	OM1	OM2	Salin1	Salin2	Bedrock	Pan	Wattab	Watsoil	Drain	Hyd	Flood	Slope
Clay1	1	0.94	0.87	0.94	-0.78	-0.92	0.97	0.94	0.89	0.83	0.94	-0.77	-0.83	0.94	0.66	0.44	0.05	0.16	0.07	0.17	-0.10	0.19	-0.09	-0.03	0.56	0.22	0.41	0.51	0.13	0.03	0.38	0.17	-0.58	0.72	-0.02	0.01
No2001		1.00	0.95	0.85	-0.84	-0.90	0.93	0.90	0.94	0.86	0.92	-0.80	-0.81	0.93	0.79	0.56	0.13	0.24	0.07	0.16	-0.10	0.02	-0.09	-0.12	0.57	0.25	0.43	0.53	0.07	0.08	0.44	0.18	-0.67	0.75	0.03	-0.11
No401			1.00	0.77	-0.79	-0.83	0.86	0.83	0.88	0.86	0.87	-0.77	-0.74	0.87	0.80	0.54	0.27	0.37	0.15	0.23	-0.14	-0.08	-0.12	-0.24	0.57	0.32	0.40	0.50	0.06	0.15	0.42	0.19	-0.68	0.77	0.04	-0.14
Shrink1				1.00	-0.57	-0.85	0.94	0.88	0.84	0.79	0.87	-0.60	-0.78	0.90	0.51	0.37	0.10	0.24	0.16	0.27	-0.08	0.20	-0.08	-0.01	0.48	0.12	0.41	0.52	0.14	-0.16	0.43	0.19	-0.47	0.58	0.04	0.02
Perm1					1.00	0.82	-0.73	-0.77	-0.74	-0.65	-0.76	0.93	0.71	-0.75	-0.83	-0.45	0.03	0.03	0.13	0.10	0.13	-0.06	0.10	0.10	-0.49	-0.33	-0.26	-0.35	-0.09	-0.37	-0.17	-0.04	0.61	-0.76	0.11	0.05
BD1						1.00	-0.90	-0.86	-0.83	-0.74	-0.88	0.78	0.86	-0.87	-0.70	-0.45	0.00	-0.10	0.03	-0.06	0.11	-0.18	0.07	0.02	-0.54	-0.25	-0.42	-0.51	-0.17	-0.07	-0.35	-0.17	0.57	-0.71	0.00	0.02
CEC1							1.00	0.93	0.89	0.84	0.93	-0.74	-0.79	0.96	0.63	0.43	0.07	0.19	0.12	0.22	-0.09	0.19	-0.08	-0.01	0.60	0.23	0.43	0.54	0.11	-0.04	0.46	0.21	-0.61	0.70	0.04	-0.01
Clay2								1.00	0.93	0.89	0.95	-0.81	-0.80	0.97	0.64	0.43	0.04	0.14	0.14	0.23	-0.11	0.17	-0.10	-0.02	0.53	0.19	0.33	0.45	0.14	-0.06	0.41	0.17	-0.56	0.72	0.04	0.06
No2002									1.00	0.93	0.92	-0.76	-0.75	0.93	0.68	0.62	0.15	0.29	0.21	0.31	-0.11	0.04	-0.09	-0.12	0.51	0.15	0.40	0.52	0.05	-0.10	0.50	0.20	-0.63	0.67	0.09	-0.12
No402										1.00	0.86	-0.73	-0.66	0.88	0.61	0.55	0.29	0.42	0.43	0.53	-0.16	0.00	-0.15	-0.24	0.53	0.22	0.35	0.47	0.06	-0.13	0.49	0.25	-0.59	0.64	0.09	-0.07
Shrink2											1.00	-0.77	-0.80	0.96	0.67	0.46	0.05	0.16	0.06	0.15	-0.13	0.14	-0.11	-0.04	0.50	0.20	0.36	0.46	0.14	0.03	0.38	0.13	-0.58	0.79	0.05	0.02
Perm2												1.00	0.69	-0.77	-0.77	-0.43	0.03	0.00	0.03	-0.02	0.15	-0.08	0.13	0.09	-0.46	-0.32	-0.27	-0.35	-0.11	-0.27	-0.21	-0.08	0.58	-0.74	0.07	0.00
BD2													1.00	-0.79	-0.61	-0.44	-0.01	-0.09	0.11	0.01	0.11	-0.14	0.08	0.05	-0.41	-0.21	-0.40	-0.47	-0.12	-0.09	-0.28	-0.09	0.50	-0.64	0.01	0.02
CEC2														1.00	0.65	0.44	0.09	0.21	0.14	0.24	-0.11	0.14	-0.10	-0.05	0.52	0.21	0.38	0.49	0.10	-0.02	0.44	0.16	-0.60	0.75	0.06	-0.03
AWC1															1.00	0.57	0.04	0.06	-0.10	-0.07	-0.09	-0.04	-0.06	-0.13	0.48	0.33	0.06	0.14	0.04	0.31	0.11	-0.09	-0.47	0.65	-0.07	-0.13
AWC2																1.00	0.08	0.19	0.05	0.13	-0.07	-0.06	-0.03	-0.09	0.42	0.28	0.21	0.28	-0.08	-0.16	0.28	0.05	-0.42	0.29	0.00	-0.20
No41																	1.00	0.92	0.75	0.70	-0.20	-0.50	-0.22	-0.65	0.02	-0.03	0.19	0.21	-0.01	0.09	0.14	0.13	-0.26	0.07	-0.01	-0.19
No101																		1.00	0.75	0.78	-0.19	-0.49	-0.22	-0.60	0.06	-0.08	0.31	0.34	-0.01	-0.11	0.28	0.22	-0.31	0.06	0.04	-0.25
No42																			1.00	0.97	-0.18	-0.27	-0.21	-0.53	0.12	0.00	0.04	0.12	-0.01	-0.24	0.20	0.20	-0.11	-0.05	0.06	-0.04
No102																				1.00	-0.19	-0.23	-0.21	-0.49	0.15	-0.02	0.11	0.20	-0.01	-0.35	0.28	0.25	-0.16	-0.03	0.08	-0.04
Inch101																					1.00	0.23	0.86	0.40	0.06	-0.08	-0.06	-0.07	-0.05	-0.13	0.09	0.20	0.04	-0.19	0.24	-0.02
Inch31																						1.00	0.21	0.65	0.12	0.04	-0.12	-0.11	0.16	-0.14	-0.06	-0.02	0.13	0.05	0.05	0.28
Inch102																							1.00	0.35	0.12	-0.08	-0.07	-0.07	-0.02	-0.15	0.06	0.17	0.05	-0.19	0.22	0.05
Inch32																								1.00	0.08	-0.08	-0.08	-0.11	0.12	-0.23	0.00	0.04	0.19	-0.14	0.16	0.22
OM1																									1.00	0.51	0.17	0.27	0.04	-0.14	0.33	0.27	-0.48	0.33	0.09	0.07
OM2																										1.00	-0.06	0.00	-0.05	0.18	-0.05	-0.10	-0.14	0.27	-0.04	0.01
Salin1																											1.00	0.94	-0.07	-0.07	0.51	0.47	-0.60	0.25	0.07	-0.23
Salin2																												1.00	-0.08	-0.11	0.58	0.45	-0.64	0.32	0.09	-0.22
Bedrock																													1.00	-0.08	-0.08	-0.01	0.05	0.15	-0.04	0.54
Pan																														1.00	-0.30	-0.29	-0.18	0.43	-0.28	-0.12
Wattab																															1.00	0.62	-0.69	0.29	0.50	-0.21
Watsoil																																1.00	-0.49	0.09	0.42	-0.10
Drain																																	1.00	-0.62	-0.20	0.26
Hyd																																		1.00	-0.05	0.03
Flood																																			1.00	-0.06
Slope																																				1.00

Step 1

All variables were entered individually in the first step to test for each variable's potential to form clusters (Appendix B2, Cluster Analysis, page B -39). Table 4 contains the variables with CCC values over 3 and indicates the level of agreement between the Average and Centroid methods. Some variables formed a relatively high number of clusters but many of the clusters contained one to a few members. Bedrock for example formed seven clusters but four of the clusters had 1, 2, 4, and 6, members. Although this stratification could contain useful information, many clusters with low membership would not have been useful in the subsequent CDA classification procedure. The Shrink1 variable, which formed 3 clusters, was chosen as the variable to include in the second step because of the consistency in results between clustering methods and because the number of clusters and CCC value had potential for growth.

Step 2

In step 2, Wattab was chosen as the next variable to include due to close agreement between the two cluster methods and again because the number of clusters and CCC value could be enlarged in the next step (Appendix B page B -75). The addition of WATTAB did not increase the number of clusters, but it did provide further discrimination between fine-textured clusters (Table 5, Step 1 vs Step 2). The variable Inch101 did have an exact match between statistics but, again, clusters contained very few members and would not have been useful in the CDA classification procedure. The problem with few members was consistent throughout the analysis for the inch10, inch3 and bedrock variables so they were not considered in subsequent steps.

Step 3

For step 3, both clustering methods indicated that Perm1 was the next significant variable to include (Appendix B page B -101). When compared to the mean values for the clusters identified in step 2, Perm1 broadened the texture clusters with respect to coarse soil conditions (Table 5). Data in Table 5 are ordered according to increasing values of the No2001 variable. The mean for the No2001 variable is intended as an aid in interpretation of the combination of permeability and shrink-swell data as related to soil texture. The No2001 variable was a clustering variable in the original analysis. It was reflective of ranges in soil texture with low percentages corresponding to coarse-textured, sandy soil and higher percentages to finer-textured, clay soil (Troiano et al., 1994). A plot of the No2001 and Perm1 variables is curvilinear in nature and indicates greater discrimination between coarse soils as permeability increases (Figure 2a). In contrast, the plot between No2001 and Shrink1 indicates a censored relationship where most of the coarse-textured soils between 20 and 40% for No2001 have shrink-swell values of zero and above 40% the shrink-swell values are positively related to No2001 values (Figure 2b). Plots based on measures of clay content (Clay1) indicate the same result (Figures 3a and 3b).

Table 4. Significant stepwise results for clustering of soil variables.

Step, and Level of Statistical Agreement and Variable(s) Entered	CCC Vaue	Number of Clusters	Comments
Step 1. Each variable tested.			
I. Exact Agreement between Average and Centroid Method			
Shrink1	3.72	3	
Pan	9.85	6	Psuedo-F not at peak value
Watttab	9.62	7	
Om2	3.13	2	Psuedo-T statistic not exact match
Bedrock	6.65	7	Many small clusters
II. Close Agreement at Same Cluster Number			
Hyd	Average-3.94	6	Psuedo-t not at Valley
	Centroid-4.22	6	
III. Close Agreement			
Flood	Average-6.31	7	
	Centroid-6.22	8	
Step 2. Shrink1 included in all tests			
I. Exact Agreement between Average and Centroid Method			
inch101	15.25	5	Only small clusters identified
II. Close Agreement at Same Cluster Number			
Wattab	Average-3.62	3	
	Centroid-3.92	3	
Perm1	Average-7.12	12	
	Centroid-9.99	12	
Bedrock	Average-6.95	12	Many small clusters
	Centroid-5.96	12	
Hyd	Average-4.58	13	
	Centroid-2.87	13	
III. Close Agreement			
Perm2	Average-6.97	3	
	Centroid-8.88	4	
Step 3 Shrink1 and Wattab included in all tests			
I. Exact Agreement between Average and Centroid Method - No exact match			
II. Close Agreement at Same Cluster Number			
Perm1	Average-4.09	7	Agreement also at 15 clusters
	Centroid-3.62	7	
Perm2	Average-4.43	13	
	Centroid-4.68	13	
III. Close Agreement			
OM2	Average-4.31	17	Peak in Psuedo-F statistic at 16
	Centrois-4.55	16	
Bedrock	Average-7.75	17	
	Centroid-3.40	16	
Step 4 Shrink1, Wattab, and Perm1 included			
I. Exact Agreement between Average and Centroid Method - No exact match			
II. Close Agreement at Same Cluster Number			
Pan	Average-8.17	19	
	Centroid-3.08	19	
Flood	Average-5.66	19	Mimicked Wattab data with only
	Centroid-3.10	19	1 observation forming a unique cluster
Watsoil	Average-6.29	14	Mimicked Wattb data with only
	Centroid-3.05	14	3 sections forming a unique cluster
OM2	Average-6.20	10	Psuedo-F not a Peak
	Centroid-5.10	10	Psuedo-F not a Peak

Table 5. Statistics for results of clustering analysis at each step.

Step, Cluster Method and Cluster Number	Number of Members	No2001		Shrink1		Wattab		Perm1		Pan	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Step 1, Average Method											
3	324	44	10	0.06	0.12						
1	84	64	11	0.74	0.19						
2	57	80	7	1.41	0.22						
Step 1, Centroid Method											
3	324	44	10	0.06	0.12						
1	84	64	11	0.74	0.19						
2	57	80	7	1.41	0.22						
Step 2, Average Method											
2	370	46	10	0.13	0.22	0.05	0.11				
1	31	79	10	1.05	0.19	0.87	0.12				
3	64	76	8	1.26	0.12	0.16	0.20				
Step 2, Centroid Method											
2	370	46	10	0.13	0.22	0.05	0.11				
1	31	79	10	1.05	0.19	0.87	0.12				
3	64	76	8	1.26	0.12	0.16	0.20				
Step 3, Average Method											
6	1	17	-	0.00	-	0.00	-	13.0	-		
7	1	26	-	0.00	-	0.73	-	10.4	-		
2	87	33	4	0.01	0.04	0.07	0.08	7.8	1.4		
5	16	41	8	0.05	0.12	0.42	0.12	5.0	1.1		
1	267	51	8	0.18	0.25	0.03	0.06	2.6	1.2		
4	51	76	8	1.29	0.27	0.09	0.13	0.6	0.3		
3	42	79	9	1.08	0.40	0.77	0.19	0.8	0.6		
Step 3, Centroid Method											
6	1	17	-	0.00	-	0.00	-	13.0	-		
7	1	26	-	0.00	-	0.73	-	10.4	-		
2	86	33	4	0.01	0.02	0.07	0.08	7.8	1.4		
5	16	41	8	0.05	0.12	0.42	0.12	5.0	1.1		
1	269	51	8	0.18	0.26	0.03	0.06	2.6	1.3		
4	50	76	8	1.30	0.27	0.08	0.13	0.6	0.3		
3	42	79	9	1.08	0.40	0.77	0.19	0.8	0.6		

Table 5 continued on next page

Table 5. Continued

Step, Cluster Method and Cluster Number	Number of Members	No2001 Mean	No2001 SD	Shrink1 Mean	Shrink1 SD	Wattab Mean	Wattab SD	Perm1 Mean	Perm1 SD	Pan Mean	Pan SD
Step 4, Average Method											
17	1	17	-	0.00	-	0.00	-	13.0	-	0.00	-
19	1	26	-	0.00	-	0.73	-	10.4	-	0.00	-
1	83	33	4	0.01	0.04	0.07	0.08	7.8	1.4	0.05	0.09
9	4	33	3	0.00	0.00	0.00	0.00	8.4	1.5	0.48	0.05
12	15	41	8	0.06	0.13	0.43	0.12	5.1	1.1	0.04	0.08
16	1	46	-	0.00	-	1.00	-	3.1	-	0.00	-
4	123	49	7	0.17	0.23	0.02	0.05	2.9	1.2	0.41	0.10
3	71	49	9	0.24	0.28	0.04	0.08	3.1	1.3	0.06	0.09
15	1	52	-	0.00	-	0.33	-	1.3	-	0.67	-
5	58	53	6	0.02	0.08	0.00	0.00	1.7	0.5	0.84	0.15
8	15	62	7	0.65	0.16	0.03	0.08	1.3	0.9	0.80	0.16
7	18	66	5	1.08	0.19	0.00	0.02	0.8	0.3	0.45	0.14
18	1	71	-	1.00	-	0.50	-	0.7	-	1.00	-
6	4	74	2	1.35	0.13	0.05	0.10	0.5	0.1	0.95	0.10
11	11	76	10	0.71	0.24	0.89	0.11	1.2	0.6	0.01	0.03
2	12	79	7	1.16	0.24	0.54	0.09	0.6	0.2	0.05	0.12
14	8	81	5	1.16	0.22	0.88	0.11	0.5	0.2	0.46	0.17
10	29	81	6	1.41	0.27	0.13	0.14	0.5	0.3	0.00	0.02
13	9	85	4	1.49	0.27	0.86	0.12	0.5	0.4	0.00	0.00
Step 4, Centroid Method											
18	1	17	-	0.00	-	0.00	-	13.0	-	0.00	-
19	1	26	-	0.00	-	0.73	-	10.4	-	0.00	-
1	104	34	2	0.00	0.02	0.06	0.08	7.2	1.9	0.08	0.12
11	16	42	8	0.07	0.14	0.42	0.12	5.0	1.1	0.05	0.08
17	1	46	-	0.00	-	1.00	-	3.1	-	0.00	-
4	157	49	7	0.13	0.21	0.02	0.06	2.7	1.1	0.47	0.15
3	43	50	7	0.25	0.23	0.05	0.07	2.9	1.3	0.03	0.07
5	27	56	4	0.04	0.09	0.00	0.00	1.3	0.1	1.00	0.00
8	15	63	7	0.65	0.16	0.03	0.08	1.3	0.9	0.80	0.16
7	20	65	5	1.05	0.20	0.00	0.02	0.8	0.4	0.44	0.13
16	1	71	-	1.00	-	0.50	-	0.7	-	1.00	-
15	1	72	-	2.00	-	0.00	-	0.1	-	0.00	-
6	4	75	2	1.35	0.13	0.05	0.10	0.5	0.1	0.95	0.10
10	11	76	10	0.71	0.24	0.89	0.11	1.2	0.5	0.01	0.03
9	34	79	9	1.28	0.33	0.13	0.14	0.7	0.6	0.00	0.02
2	12	79	7	1.16	0.24	0.54	0.09	0.6	0.2	0.05	0.12
14	2	80	6	1.36	0.20	1.00	0.00	0.5	0.1	0.73	0.08
13	6	80	6	1.09	0.19	0.84	0.10	0.5	0.2	0.38	0.07
12	9	85	4	1.49	0.27	0.86	0.12	0.5	0.4	0.00	0.00

Figure 2. a) Plot of sectional estimates for permeability (Perm1) vs No 200 sieve (No2001) of the surface layer; b) sectional estimates for shrink-shwell potential (Shrink1) vs No2001.

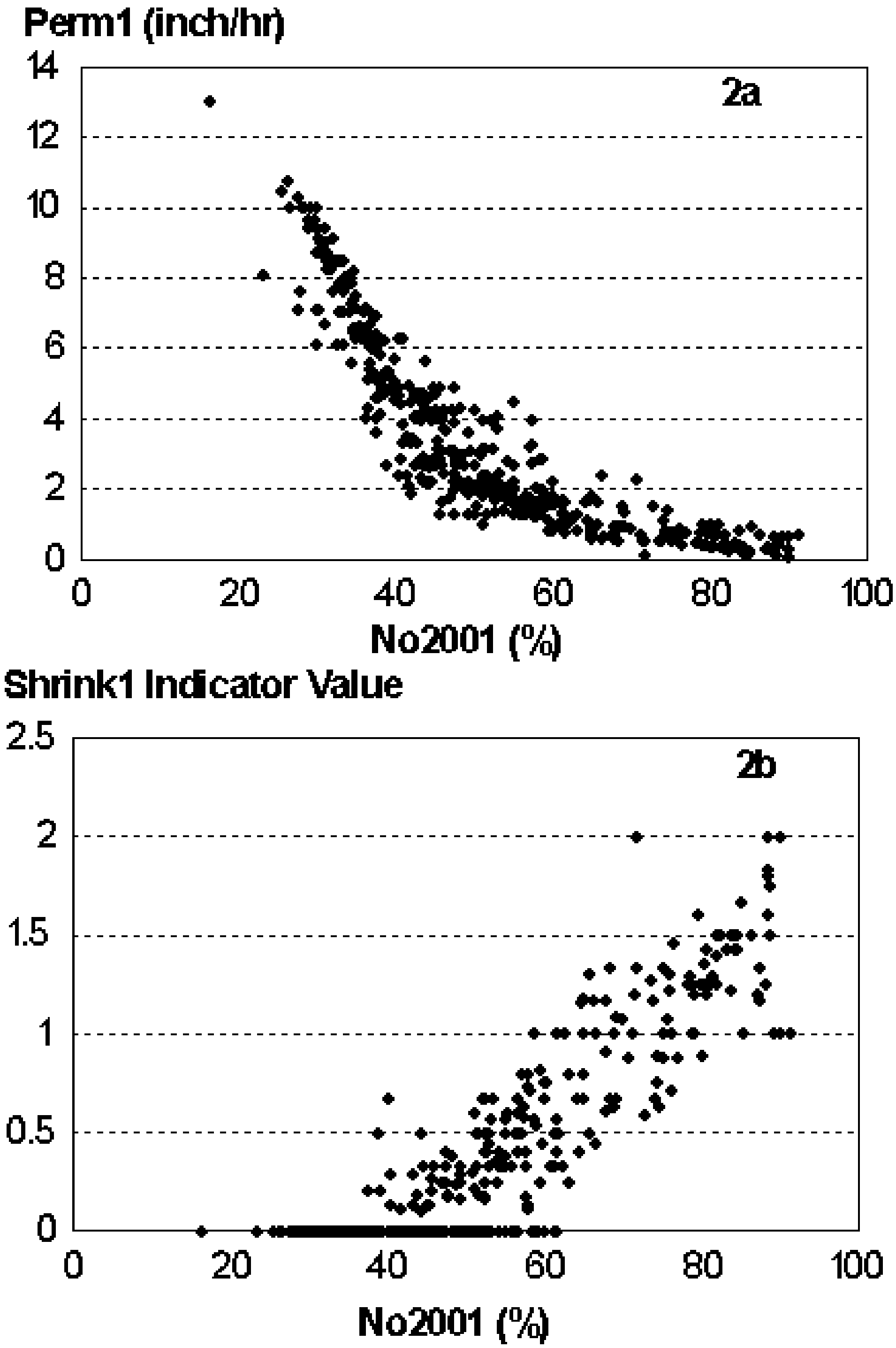
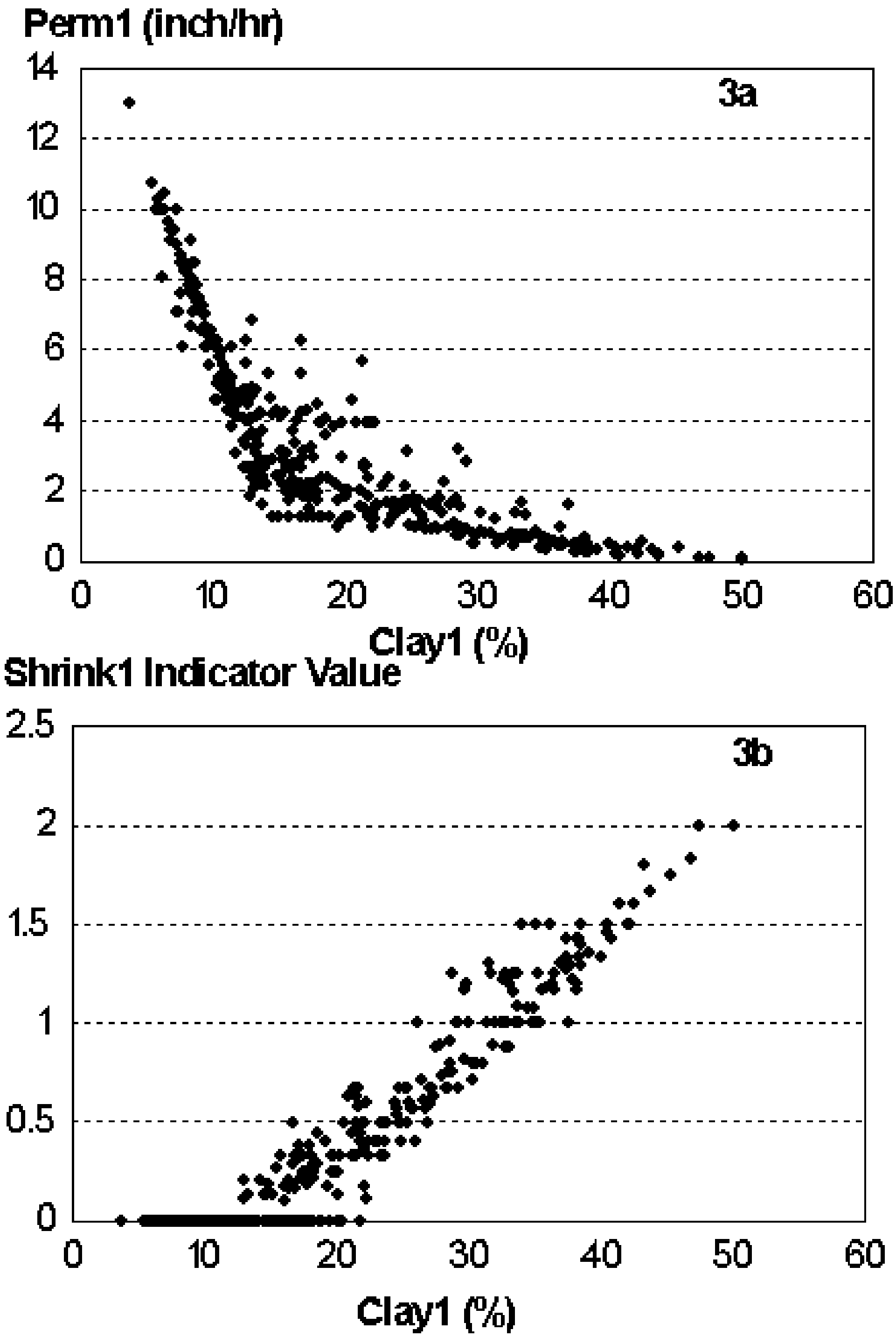


Figure 3. a) Plot of sectional estimates for permeability (Perm1) vs clay content (Clay1) of the surface layer; b) sectional estimates for shrink-shwell potential (Shrink1) vs Clay1.



An interesting aspect of the distribution of sections between clusters formed with the combination of Shrink1, Wattab, and Perm1 is that only approximately 20 percent of the sections were partitioned into a coarse-textured cluster (Table 5). Thus, potentially 80% of the detections occurred under soil conditions where leaching with simple percolation may not have been the predominant pathway for movement to ground water.

Step 4
In the fourth step, no exact results were indicated between methods, but the results for inclusion of the Pan variable indicated a similar number of clusters formed at 19 (Appendix B page B-126). The CCC values for this four-variable combination were above 3.0, and a peak in the Psuedo-F and valley in the Psuedo-t^2 statistics were observed for both clustering method (Tables 4 and 5). A descriptive comparison of the mean values for the clusters indicated similar cluster formation (Table 6). The shaded cells in Table 6 are clusters that were an exact match between methods, which tended to be clusters with lesser members. Although there was general agreement in the distribution of the other clusters, the exact partitioning of members between other clusters differed. The Average method tended to provide greater discrimination between coarser-textured soils whereas the Centroid method tended to provide greater discrimination between the finer-textured soils.

Owing to the greater degree of divergence creeping into the analysis and the increasing difficulty in interpretation, the usefulness of results at further steps was questioned and the process ceased at four variables.

Profiling Algorithm
The result from the Average clustering method was used in the CDA profiling analysis. The Average method was chosen based on the original study results (Troiano et al. 1997) and on the increasing value observed in the CCC value with each succeeding step in the cluster analysis. The CDA analysis was conducted on 12 of the 19 clusters which was 452 of the 465 KC sections; those with 4 or fewer members were not included (Table 5). In the previous analysis reported in 1997 (Troiano et al., 1997), the first 2 CVs accounted for 98% of the variation in the original 254 KC sections and they were sufficient in defining the location of the clusters. For this revised analysis, the first 2 CVs accounted for 79% of the total variation and 3 CVs accounted for 95% of the total variation (Appendix 3a, CDA Analysis Results page B-149). With respect to the classification algorithm, a section was considered a member of a KC soil cluster if the Euclidean distance between the observation and the cluster centroid was less than or equal to the radius of either the circular or spherical 95% population tolerance interval (Appendix 3b, CV means for raw data page B-157).

Results for CDA classification algorithm of the KC sections were compared between algorithms based on 2 and then 3 CVs. For 2 CVs, 82 of the 452 sections were not classified into the cluster that was previously determined by the Average cluster analysis (Appendix 3c page B-158). Of the 82 sections, only 3 sections did not fall within the circular population tolerance interval for any of the soil clusters. The remaining 79 sections were

Table 6. Descriptive comparision of cluster formation between Average and Centroid clustering results at the 19 cluster solution.

Predominant Cluster Characteristics	Average Method		Centroid Method	
	Cluster Designation	Number of Members	Cluster Designation	Number of Members
A. Very Coarse-Textured	17	1	18	1
B. Very Coarse-Textured + Water Table	19	1	19	1
C. Coase-Textured	1	83	1	104
D. Coarse-Textured + Pan	9	4	-	-
E. Medium-Coarse-Textured + Water Table	12	15	11	16
F. Medium-Coarse-Textured + Water Table	16	1	17	1
G. Medium-Coarse-Textured + Pan	4	123	4	157
H. Medium-Coarse-Textured	3	71	3	43
I. Medium-Textured + Pan + Water Table	15	1	-	-
J. Medium-Textured + Pan	5	58	5	27
K. Medium-Fine-Textured + Pan	8	15	8	15
L. Medium-Fine-Textured + Pan	7	18	7	20
M. Medium-Fine-Textured + Pan + Water Table	18	1	16	1
N. Medium-Fine-Textured + Extreme Shrink1 Value	-	-	15	1
O. Fine-Textured + Pan	6	4	6	4
P. Fine-Textured + Water Table	11	11	11	11
Q. Fine-Textured	10	29	9	34
R. Fine-Textured + Water Table	2	12	2	12
S. Fine-Textured + Water Table + Pan	14	8	14	2
T. Fine-Textured + Water Table + Pan	-	-	13	6
U. Very Fine-Textured	13	9	13	12

miss-classified between soil clusters. For 3 CVs, 44 sections were not classified into the cluster that was previously determined by cluster analysis (Appendix 3d page B-167). Of the 44 sections, fifteen sections did not fall within the spherical population tolerance interval for any of the soil clusters. The overall error rate for misclassification decreased from 17.6% to 9.5% for 2 to 3 CVs, respectively, but the proportion of sections not classified into a soil cluster increased from 0.6 to 3.2%, respectively. Apparently, the 3rd CV provided greater graphical separation of the clusters, as indicated by the reduction in the overall rate of misclassification. Owing to the large number of clusters formed, some misclassification between clusters would be expected, especially between those located between the extremes of the axes. Potential overlap of clusters is illustrated in a plot of the 2-dimensional circular tolerance intervals for CV1 and CV2 (Figure 4). It is interesting that the percentage of sections not classified into any of the soil clusters, was at 3.2%, which was within the 5% spherical population tolerance interval.

Comparison of CALVUL Sectional Estimates with Digitized Soil Data

The 3-CV profiling algorithm was applied to soil data developed for Fresno and Tulare Counties (Figure 5). The geographic pattern mimicked the general soil description for these soil survey areas. A test of the accuracy of the sectional CALVUL estimates was conducted for the hardpan soil condition whereby they were compared to sectional estimates derived from digitized data for the Central Tulare County Soil Survey. This data is one of the few digitized soil surveys now available through the NRCS. Soil MUIDS with a hardpan indicator were numbered 660124, 660125, 660145, 660154, 660155, and 660159 in the Central Tulare Survey. For CALVUL estimates, sections defined as containing a hardpan were from soil clusters 4, 5, 7, and 8 in Table 5, which were used in the CDA profile algorithm (Table 5, Step 4 - Average method).

In Figure 6, the surface areas covered by soil MUIDs with hardpan are indicated in solid blue upon the gray background. The CALVUL estimates for sections containing hardpan soil are illustrated as the dark blue outlined squares. Good spatial correlation is indicated by the overlap of these two data sets. Some lack of correspondence between the data sets was measured when the sectional CALVUL hardpan values were regressed and plotted against the values determined from the digitized soil database (Figure 7). Sections with soils containing values of the Slopeh variable > 15% were removed from this analysis in order to minimize bias in the CALVUL estimates, which had excluded these MUIDS. In evaluating the regression, a few sections at digitized values of 0% had an indication of hardpan in the CALVUL estimate and conversely, at digitized values of 100 a few CALVUL sections had indication of soils other than hardpan. Please note that data at 0,0 and 1,1 co-ordinates are represented by a single point, whereas there were many points at these co-ordinates. A source of error in the digitized NRCS data set was observed when the digitized maps were compared to the hard copy maps; some digitized polygons were mislabeled. For example, a digitized section might indicate no soil with a hardpan when, in actuality, a small polygon was present on the hard copy. Even with this source of error, a comparison of the scatter plot to the 1:1 line indicated that the CALVUL model values overestimated the values at the low end of the scale and underestimated the values at the high end. Overestimation by the CALVUL estimates at the low end indicate a conservative effect where vulnerable acreage would be overestimated, an effect caused

Figure 4. Plot of the first 2 canonical variates obtained from a Canonical Discriminant Analysis of the ouput from step 4 clustering by the Average Method. Numbers in circles are clusters as indicated in Table 5.

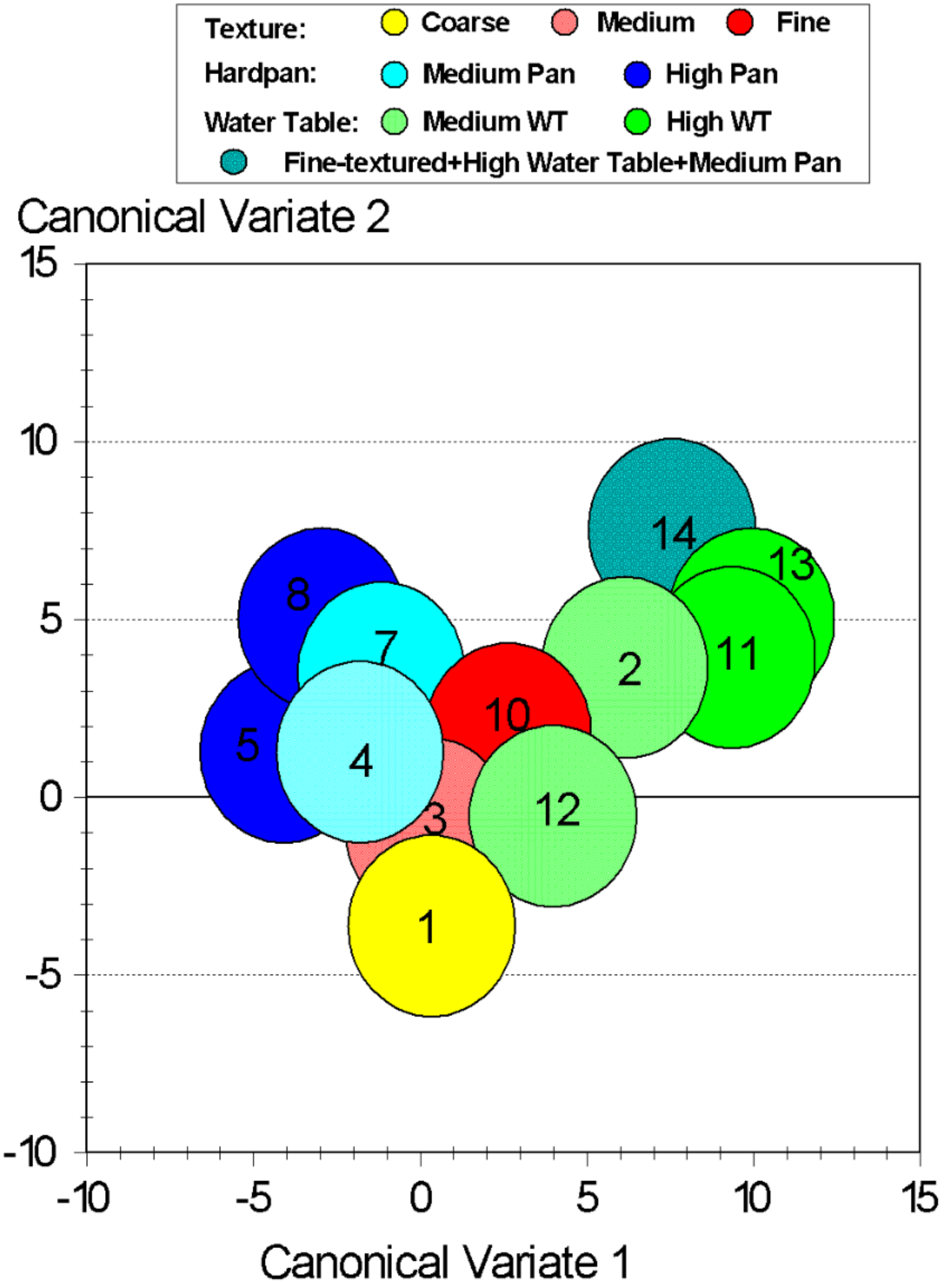


Figure 5. Map of all CALVUL clusters characterized in Fresno and Tulare Counties.

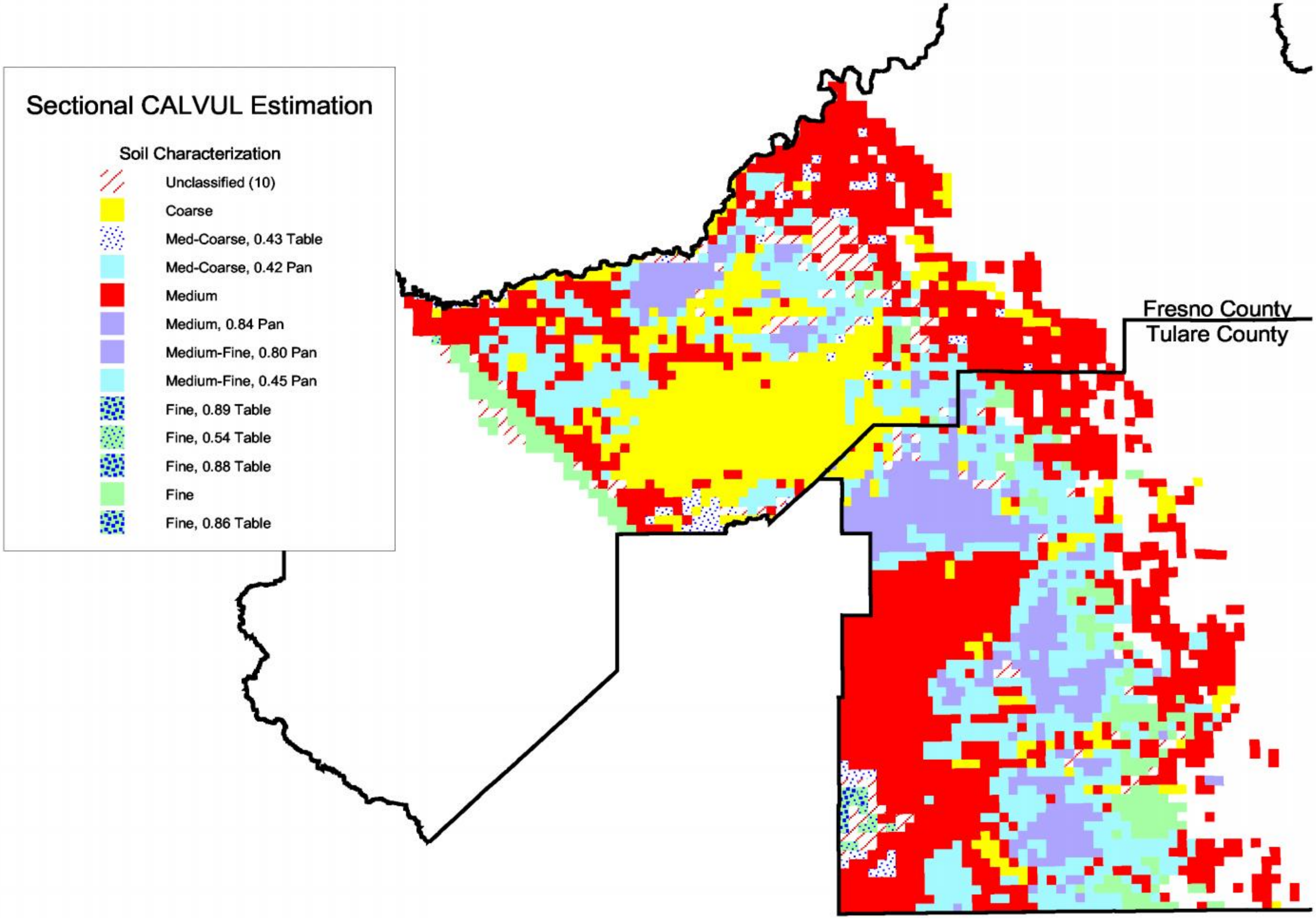


Figure 6. Overlay of CALVUL Model estimates for sections characterized as runoff on NRCS Central Tulare County Soil Series 660 with pan soils.

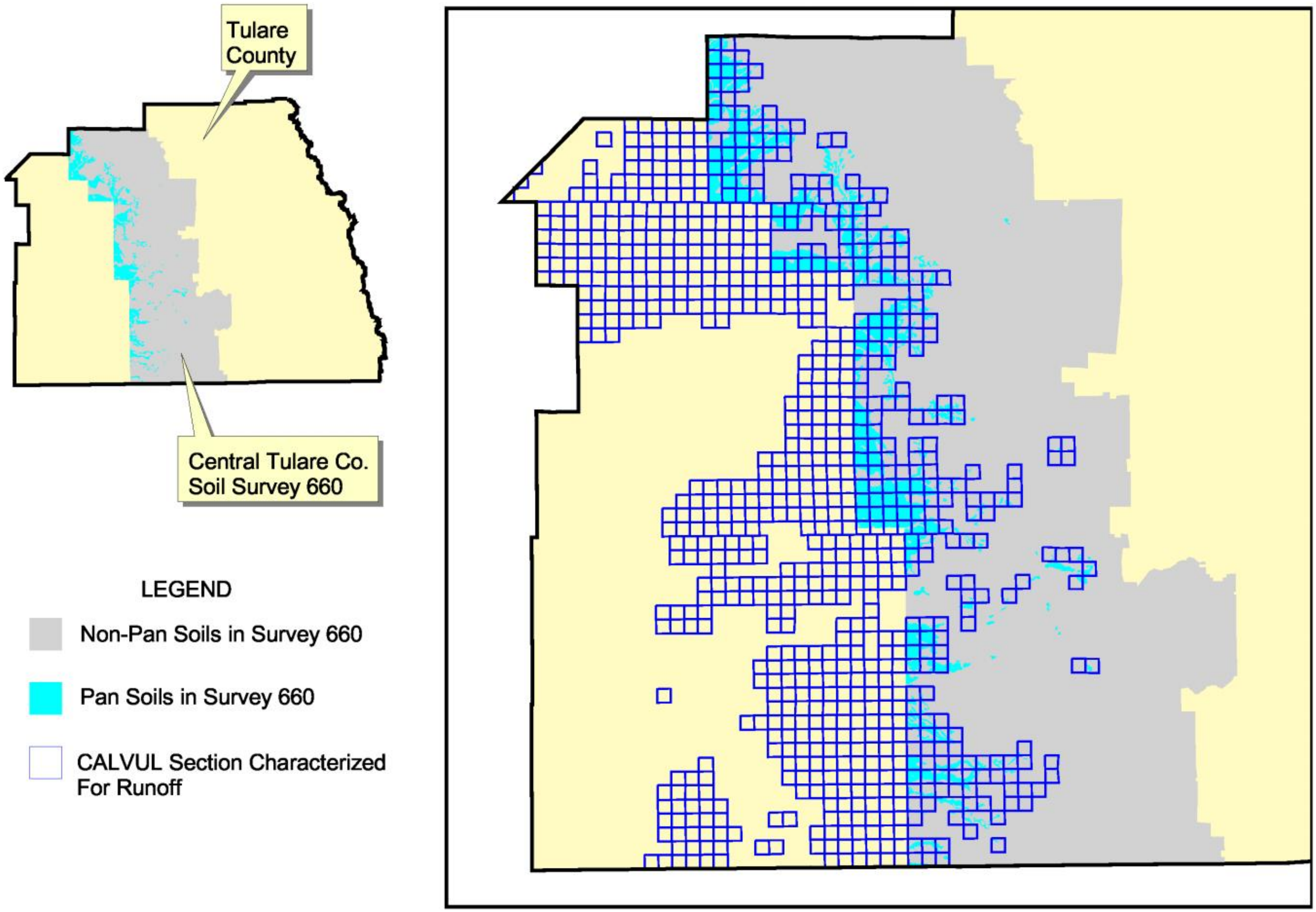
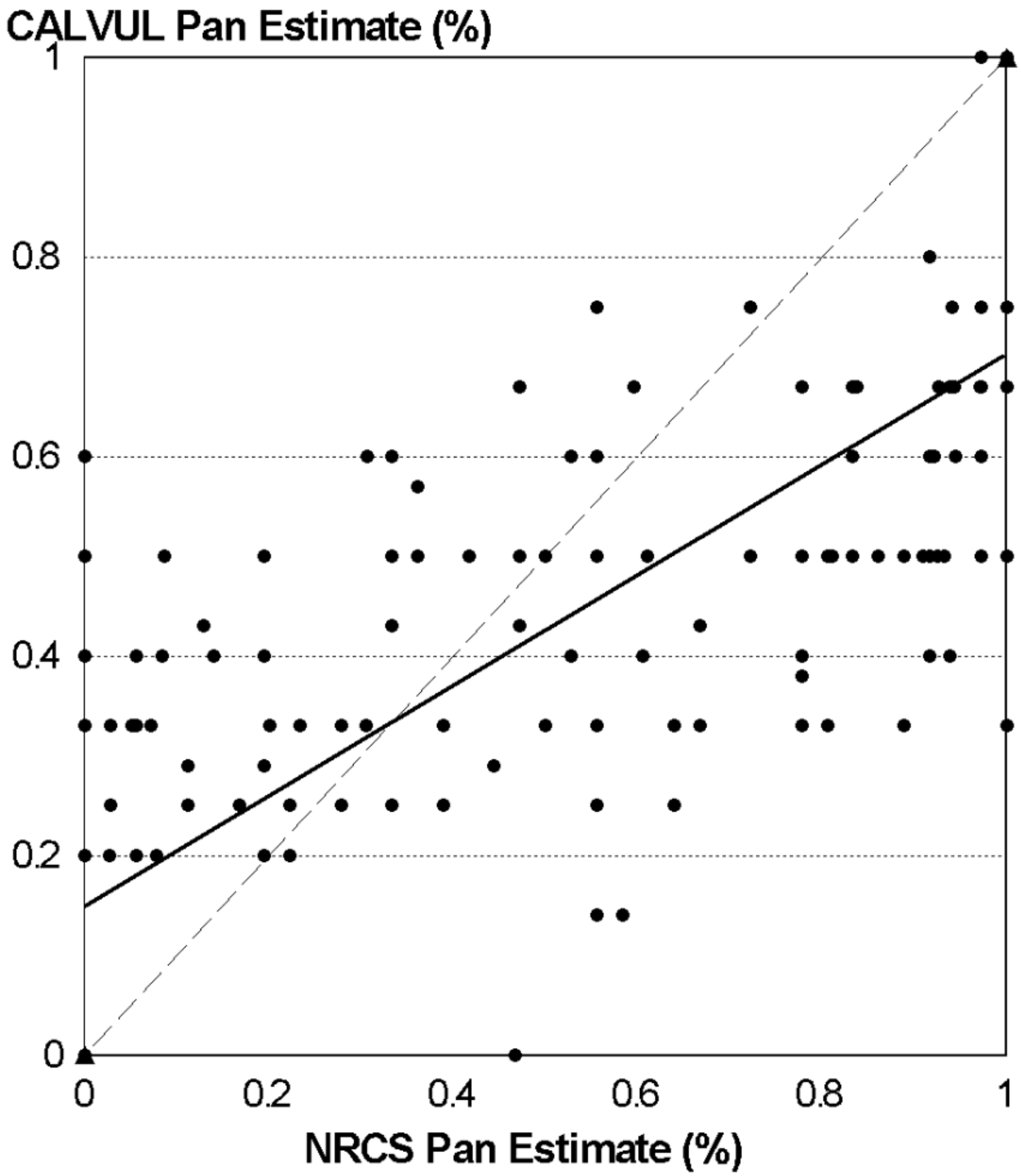


Figure 7. Estimates of percentage of hardpan soils in a section compared between CALVUL and digitized NRCS soil data. Solid line is trend line and Dashed line is the 1:1 fit.

ANOVA:	Source	DF	Mean Square	F
	Model	1	7.0803	217.82
	Error	156	0.0325	

Regression:	Variable	Parameter	Std Error	Prob>T
	Intercept	0.1503	0.0236	0.0001
	NRCSpan	0.5527	0.0374	0.0001



by the averaging procedure used to produce sectional values. This limitation will be recognized during implementation of the model results by stressing that the CALVUL results are estimates and that, when in doubt, they should be used in conjunction with either published maps or digitized data.

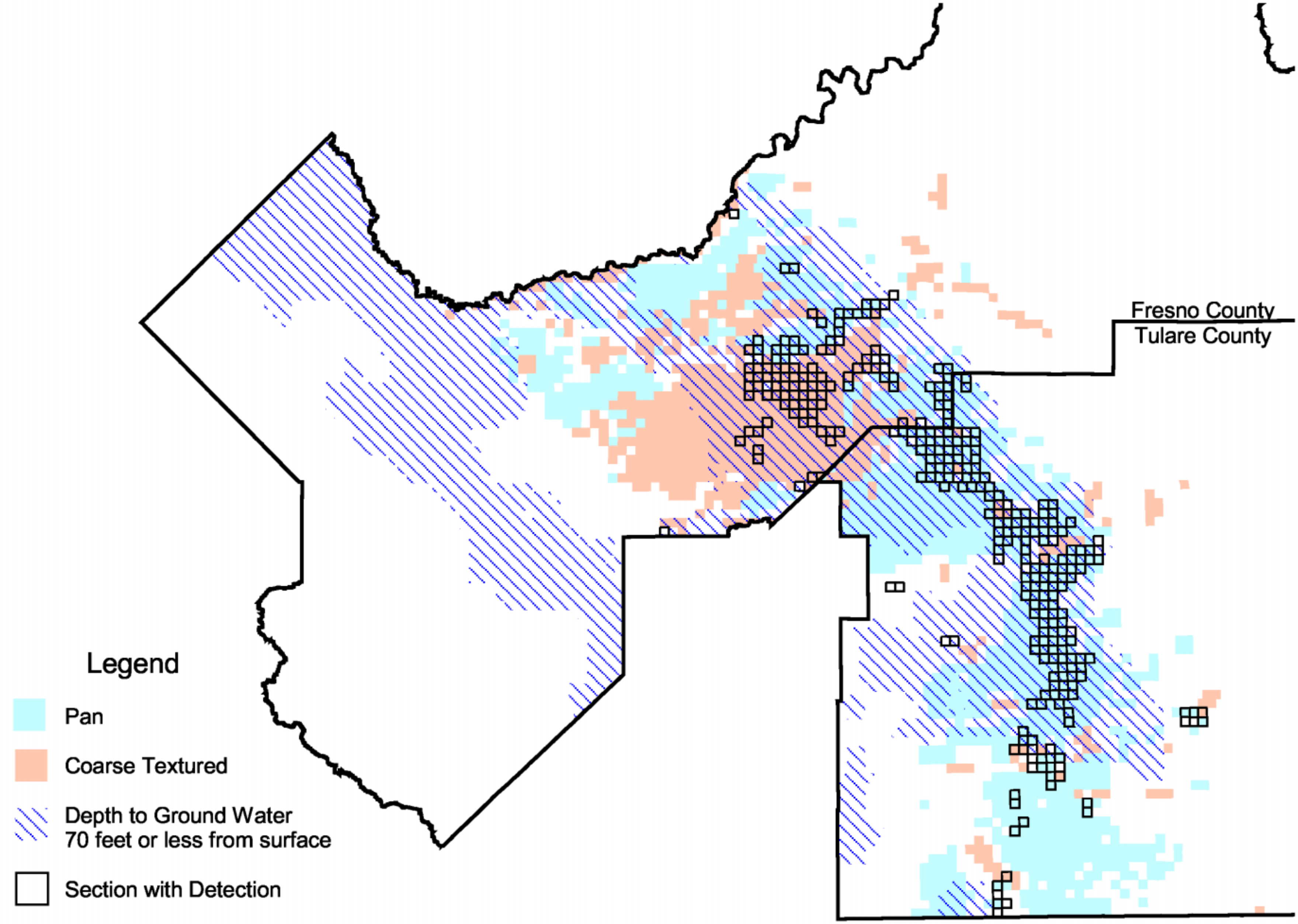
Application of the CALVUL Approach to California's Ground Water Protection Program

The cluster assignment for any section of land only represents a potential for that land area to be associated with a vulnerable condition. The next step is to link the vulnerable soil condition with a pathway or pathways for pesticide movement to ground water. This data has been developed for two of the vulnerable soil conditions. For the coarse soil condition, leaching with simple percolation from the site of application has been identified as the predominant pathway for movement of residues to ground water (Troiano et al., 1993). Consequently, effective irrigation management has been identified as the method to mitigate movement of residues. Coarse soil areas are predominantly located in Fresno County and they are shaded in yellow (Figure 5).

Movement of pesticide residues in runoff water has been identified as a pathway for pesticide movement to ground water in hardpan soils. Sections with hardpan soil are denoted in the various shades of blue in Figure 5. Hardpan soils are predominant in Tulare County but they are also present in Fresno County, primarily along the eastern side of the Central Valley. Investigations conducted in this hardpan soil area have demonstrated widespread contamination of ground water caused by movement of winter rain runoff water that contains pre-emergence herbicide residue into dry wells or into areas with high infiltration rates (Troiano and Segawa, 1987; Braun and Hawkins, 1991). Runoff-prone soils have a low infiltration rate so mitigation measures are different than for leaching-prone soils. Pre-emergence herbicides are usually broadcast onto the soil surface and rainfall is suggested as the method to incorporate the broadcast residues into the soil matrix. But for runoff-prone soils, rainfall should not be a suggested method of incorporation. Instead, residues should be mixed or moved into the soil prior to exposure to winter rainfall by some other method of incorporation such as mechanical incorporation. Mechanical incorporation has been shown to greatly reduce the mass of simazine carried off the field in simulated-rain runoff applied to citrus row middles (Troiano and Garretson, 1998).

A large portion of each county would be classified as vulnerable if the vulnerability analysis relied solely upon soil data. Soil data, however, is not the only piece of geographic information that will be used to identify vulnerable areas. Based on data developed for the norflurazon retrospective well study, depth-to-ground water data have been developed as another layer used to identify areas with higher pollution potential. The procedure used to determine a sectional average depth of 70 feet or shallower as the cut-off for areas with higher potentials for ground water contamination was described in Troiano et al., (1999). Sections in Fresno and Tulare Counties with spring average DGW less than 70 feet are indicated as the lined areas in Figure 8. The intersection of DGW data overlain with the sectional data for coarse or hardpan sections provides the

Figure 8. Sections in Fresno and Tulare Counties characterized as having pan or coarse soils, a depth to ground water of 70 feet or less from the surface, and pesticide detections.



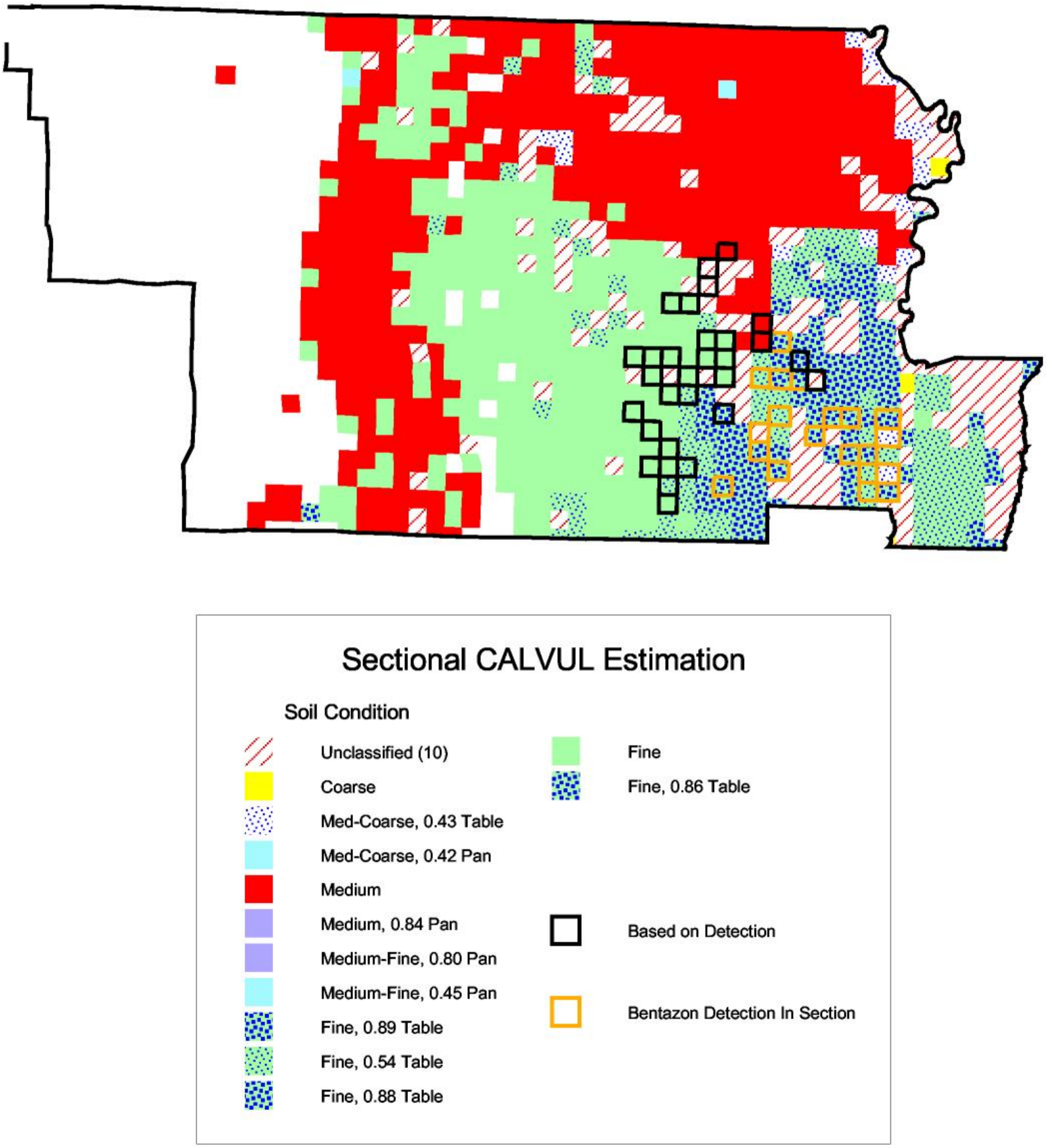
geographical identification for sections of land with a high probability for pesticide movement and subsequent detection in ground water (Figure 8). These areas will be designated as Ground Water Protection Areas (GWPAs) and they are areas where processes of movement to ground water have been identified and investigations into mitigation measures have been conducted. The dark outlined squares (sections) in Figure 8 are the KC sections where pesticides have been detected in well water. A great majority of the KC sections are contained within the area described as highly vulnerable. The KC sections that fall out of the GWPAs will require further investigation but they will also be regulated by assignment into the leaching or runoff category.

For comparative purposes, the 3-CV profiling algorithm was also applied to data developed for Glenn County (Figure 9). Glenn County contains mostly soils in the medium (red) and fine-textured (green) soil clusters. Inclusion of the indicator for a seasonal water table provides an interesting division between the finer-textured clay soils. Residues have been detected in fine-texture soil with and without a high water table during the winter months. But, detection of bentazon residue in this area has been confined to fine-textured soil with a seasonal water table, which are sections outlined in gold in Figure 9. Since bentazon detection was associated only with rice production, is the association with this soil condition merely due to the location of rice paddys? Alternatively, does the seasonal water table have an effect that exacerbates the movement of bentazon to ground water as compared to the area with fine-textured soil that does not have a seasonal water table? These questions exemplify how the CALVUL model can aid in the investigation of local processes by which pesticides move to ground water. If the seasonal water table soil feature provides some insight into pesticide movement in this area, it may also prove to be an important factor in the development of appropriate mitigation measures.

Summary

We are proposing to use the geographical identification of highly vulnerable areas, denoted as Ground Water Protection Areas (GWPAs), as the basis for proposed regulations where mitigation measures will be implemented to prevent further movement to ground water. Currently, GWPAs are identified as sections in coarse or hardpan soil clusters that have sectional estimates of DGW at 70 feet or less. As indicated in Figure 8, this is an area where numerous wells have been shown to contain pesticide residues. The regulations will also apply in areas where residues have not yet been detected in well water but where soil and DGW indicate a similar potential for contamination. The application of the CALVUL results will enable DPR to focus resources on further demonstration and implementation of mitigation measures in these areas. In addition, the CALVUL model results will also be used in investigations into processes of contamination in other vulnerable soil conditions, such as the clay soil conditions noted in Glenn County (Figure 9).

Figure 9. Map of all CALVUL clusters characterized in Glenn County.



References

Aldenderfer, M.S. and Blashfield, R.K.: 1984, *Cluster Analysis.* Sage University Paper series no. 07-044. Sage Publications, Inc., Berverly Hills, CA.

Bergstrom, L., A. McGibbon, S. Day, and M. Snel. 1991. Leaching potential and decomposition of clopyralid in Swedish soils under field conditions. Environ. Toxicol. and Chem. 10:563-571.

Braun, A.L., and L.S. Hawkins. 1991. Presence of Bromacil, Diuron, and Simazi ne in Surface Water Runoff from Agricultural Fields and Non -crop Sites in Tulare County, California, Environmental Monitoring and Pest Management Branch, Department of Food and Agriculture, Sacramento, CA. PM 91 -1.

Burlinson, N.E., L.A. Lee, and D.H. Rose nblatt. 1982. Kinetics and products of hydrolysis of 1,2-dibromo-3-chloropropane. Environ. Sci. Technol. 16:627 -632.

Chen, C., D.M. Thomas, R.E. Green, and R.J. Wagenet 1993. Two -domain estimation of hydraulic properties in macropore soils. Soil Sci. Soc. Am. J. 57:680-686.

Connelly, L. 1986. AB2021-Pesticide Contamination Prevention Act. Article 15, Chapter 2, Revision 7, Food and Agricultural Code, California.

Davis, R.E., and F.F. Foote. 1966. 'Chapter 23', Surveying theory and practice. Fifth edition, New York, N.Y.

Fowkles, E.B, Gnanadesikan, R. and Kettenring, J.R.: 1988, 'Variable Selection in Clustering', J. of Classification, **5,** 205-228.

Graham, R.C., A.L. Ulery, R.H. Neal, and R.R. Teso. 1992. Herbicide residue distributions in relation to soi l morphology in two California vertisols. Soil Science 153:115-121.

Hallberg, G.R. 1989. Pesticide pollution of groundwater in the humid United States. Agriculture, Ecosystems and Environment 26:299 -367.

Holden, L.R., J.A. Graham, R.W. Whitmore, W.J. Ale xander, R.W. Pratt, S.K. Liddle, and L.L. Piper. 1992. Results of the national alachlor well water survey. Environ. Sci. and Technol. 26:935-943.

Johnson, B. 1991. Setting Revised Specific Numerical Values. April, 1991 Pursuant to the Pesticide Contamin ation Prevention Act. Environmental Hazards Assessment Program, California Department of Pesticide Regulation, Sacramento, CA. EH 91 -6.

Kalinski, R.J., W.W. Kelly, I. Bogardi, R.L. Ehrman, and P.D. Yamamoto. 1994. Correlation between drastic vulnerabilities and incidents of VOC contamination of municipal wells in Nebraska. Ground Water 32:31-34.

Maes, C. M., Pepple, M., Troiano, J.,Weaver, D., Kimaru, W., and SWRCB Staff. 1992. Sampling for Pesticide Residues in California Well Water: 1992 Well Inventory Data Base, Cumulative Report 1986-1992,. Environmental Hazards Assessment Program, California Department of Pesticide Regulation, Sacramento, CA. EH 93-02.

National Research Council. 1993, Groundwater Vulnerability Assessment: Predicting Relative Contamination Potential Under Conditions of Uncertainty. Water Science and Technology Board, National Research Council, National Academy Press, Washington D.C.

Roux, P.H., R.L. Hall, and R.H. Ross Jr. 1991. Small-Scale retrospective groundwater monitoring study for simazine in different hydrogeological settings. Groundwater Monit. Rev. XI: 173-181.

SAS Institute Inc: 1983, *Cubic Clustering Criterion*. *SAS*Technical Report A-108*, SAS Institute Inc., Cary, NC.

SAS Institute Inc.: 1988, *SAS/STAT User's Guide, Release 6.03 Edition.* SAS Institute Inc., Cary, N.C.

Soil Conservation Service. 1975. Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. Soil Conservation Service, U.S. Department of Agriculture, U.S. Govt. Print. Off., Washington D.C.

Soil Conservation Service. 1983. National soils handbook. Soil Conservation Service, U.S. Department of Agriculture, U.S. Govt. Print. Off., Washington D.C.

Soil Survey Staff. 1997. Title 430 National Soil Survey Handbook, Revision Issued December, 1997. National Resources Conservation Service, Washington D.C., U.S. Government Printing Office, December 1997.

Troiano, J., and R.T. Segawa. 1987. Survey for Herbicides in Well Water in Tulare County, January 1987. Environmental Hazards Assessment Program, California Department of Pesticide Regulation, Sacramento, CA. EH 87-01.

Troiano, J., and C. Garretson. 1998. Movement of simazine in runoff water from citrus orchard row middles as affected by mechanical incorporation. Journal of Environmental Quality. 28:488-494.

Troiano, J., J. Marade, and F. Spurlock. 1999. Empirical modeling of spatial vulnerability applied to a norflurazon retrospective well study in California. Journal of Environmental Quality 28:397-403.

Troiano, J., Johnson, B.R., Powell, S. and S. Schoenig. 1994, Use of cluster and principal component analyses to profile areas in California where ground water has been contaminated by Pesticides. Environmental Monitoring and Assessment 32:269 -288.

Troiano, J., C. Nordmark, T. Barry, and B. Johnson. 1997. Profiling areas of ground water contamination by pesticides in California: Phase II - evaluation and modification of a statistical model. Environmental Monitoring and Assessment 45:301 -318.

Troiano, J, C. Garretson, C. Krauter, J. Brownell, and J. Hutson. 1993. Influence of amount and method of irrigation water application on leaching of atrazine. Journal of Environmental Quality, 22:290-298.

U.S. Environmental Protection Agency. 1992. ANOTHER LOOK: National Pesticide Survey Phase II Report, EPA 570/9-91-020, Jan., 1992. Office of Water, USEPA Washington D.C.