PROFILING AREAS VULNERABLE TO GROUND WATER CONTAMINATION BY PESTICIDES IN CALIFORNIA

FINAL REPORT TO THE U.S. ENVIRONMENTAL PROTECTION AGENCY
FOR CONTRACT # E-009565-01-0

Prepared by:

John Troiano, Bruce Johnson, Sally Powell, and Steve Schoenig

California Environmental Protection Agency
Department of Pesticide Regulation
Environmental Monitoring and Pest Management Branch
1220 N Street, Room A-149
Sacramento, California 95814

ACKNOWLEDGEMENT

This study was made possible through the assistance of USDA Soil Conservation Service staff. We gratefully thank Carole Jett, Dave Smith, and Kit Paris, soil scientists with the Soil Conservation Service at Davis, California for their assistance in obtaining the MUIR soils database and for their assistance in interpreting the information it contains. We also thank DPR Senior Research Scientist Bob Teso for supplying the soil mapping unit data base for vulnerable soils which was produced under contract with Tom Rice, soil scientist at the California Polytechnic State University San Luis Obispo. Lastly, funding by the US-EPA provided the impetus for this project and we thank Tom Broz EPA region IX San Francisco, California for his support and enthusiasm for the study.

ABSTRACT

Identifying areas vulnerable to ground water contamination by pesticides is desirable because pollution prevention policies could be developed for specific locations. Previous attempts to correlate predicted levels of vulnerability with measures of the absence and/or presence of pesticide residues in well water have not been entirely satisfactory. Poor correlation between predicted level of vulnerability and occurrence of pesticide residues in well water may have been caused by assuming that only the leaching pathway was involved or by uncertainties in the use of well sampling data as an indication of vulnerability. An alternative approach was devised that produced classification algorithms based on climatic and soil data from known vulnerable (KV) sections. KV sections in California are defined as 1 square mile areas of land where pesticide residue has been detected in well water samples and the detection attributed to nonpoint source agricultural applications. Clustering procedures were used to group similar KV sections first with respect to climate data and then with respect to soil data. Principal Components Analysis was used to construct soil profiles of the clusters. The profiles were used as the basis for a classification procedure to determine if soil properties of candidate sections with unknown vulnerability were similar to profiles developed for KV sections. Since this scheme is based only on data from KV sections, candidate sections with dissimilar profiles cannot be considered as non-vulnerable; they receive a status of non-classifiable. However, the process is flexible and it can revised to incorporate updated well sampling information.

INTRODUCTION

Identification of areas vulnerable to ground water contamination by pesticides is desirable because pollution prevention policies could be developed for specific locations. One approach to identifying vulnerability has been to: 1) devise a vulnerability index based on variables thought important in facilitating pesticide movement to ground water, usually assuming the leaching pathway; 2) stratify land areas based on the vulnerability index; 3) obtain data on the detection of pesticide residue in well water; and 4) use percentage of detections as a discriminator variable in analyses conducted to test correspondence with the vulnerability index.

Tests conducted with indices derived from the DRASTIC model are an example of this approach. In DRASTIC, indices of vulnerability are derived from a series of weights and ratings of seven hydrogeologic variables which experts agreed were important determinants in leaching of pesticides to ground water (Aller et. al., 1985). The correspondence between the detection of pesticide residue in well water and DRASTIC indices, generated for county-wide areas, have been statistically tested in three studies (EPA, 1992; Balu and Paulsen, 1991; Holden et. al., 1992). None of the studies found a good correspondence between occurrence of residues and the DRASTIC scores.

Problems related to well sampling may have caused unfavorable results with this approach. First, presence of pesticide residue in well water may not solely result from leaching through soil via the normal route of water percolation. Observations of construction and quality of a well are usually made during a study to ensure that local streaming from the surface to ground water had not occurred, but movement to ground water may occur through other pathways that are difficult to investigate. For example, collection of runoff water from rainfall or irrigation may be shunted to special drainage wells or to fast draining areas of soil. Contamination could then result from an unexpected route of water movement through the soil. Second, the probability of detecting pesticides in well water is complicated by the location of the well in relation to depth and direction of ground water flow from contaminated areas. For example, a domestic well situated near and downstream (in terms of ground water flow) of an agricultural field would appear to be a good candidate sampling site because it should reflect local conditions. However, residue that has leached from the nearby field may encounter ground water at a stratum above that tapped by the well causing the residue to bypass the well. Determining specifics of ground water flow and well location for each sampled well is usually not feasible when conducting large-scale field studies.

Wilkersen et. al., (1985) used an empirical discriminant analysis approach to produce a classification equation for vulnerability. Their approach was to: 1) identify land use, geographic, and well construction variables for 1-square-mile areas designated as sections in the Public Land Survey System (PLSS) (Davis and Foote, 1966); 2) derive a classification variable for vulnerable and non-vulnerable areas which was the presence or absence of pesticide residues in well water sampled in a section; and 3) use discriminant analysis to a produce classification equation for vulnerability. A discriminant classification model was developed from data for 10 sections that were selected from 3 adjacent townships in an agricultural region of Fresno County. DBCP had been detected in 7 sections which were classified as vulnerable. The remaining 3 sections were identified as non-vulnerable.

The discriminant model which contained 4 variables correctly classified the original 10 sections. However, when tested against an independent data set, sections with DBCP detections were correctly classified as vulnerable whereas non-vulnerable sections without DBCP detections were misclassified as vulnerable. Well sampling data available after that study indicated that nearly all sections in the test townships now contain positive DBCP detections. If the study had been conducted at a later date, the entire area would have been classified as vulnerable (Brown et. al., 1986) and a discriminant analysis would not have been possible. As illustrated by this example, the dynamic nature of well sampling evidence should be considered when ratios of the presence or absence of residues in well water are used as classification variables for vulnerability.

In a similar discriminant procedure employed by Teso et. al. (1988), soil data were used to develop a discriminant function for the occurrence of either DBCP contaminated or uncontaminated sections in Fresno county. When tested against an independent data set describing DBCP contamination in Merced county, a 40% misclassification rate was measured.

Since the attempts to devise classification systems that predict levels of vulnerability have not been entirely satisfactory, an approach was devised that profiled known vulnerable (KV) sections in California. Cluster analysis was first used to identify groups of KV sections with similar climate and soil conditions. Then, a classification algorithm was derived to determine whether soil conditions of new candidate sections of unknown vulnerability matched KV section profiles. This type of approach has been described as a Hydrogeologic Setting Comparison (HSC) where areas are judged similar based on hydrogeologic character (Marshall, 1991). Previous HSC efforts utilized a restricted set of hydrogeologic variables that were interpreted with respect to the leaching pathway (Kisel et. al., 1982; Fisher and Reid, 1986; Sacha et. al., 1987). This current work expanded upon the HSC approach in six ways: 1) the number of climatic and soil variables considered as identification variables was increased; 2) data were obtained that could be resolved at the section level, a 1 square-mile area; 3) no assumptions were made about the causes of ground water contamination because, according to our experience, leaching is only one of several possible causes of ground water contamination from nonpoint source pollution; 4) clustering techniques were used to chose combinations of climate and soil variables that formed unique clusters of vulnerable sections; 5) classification algorithms were developed from the clustering results; and 6) the entire process could be revised to accommodate new information on vulnerable areas when a greater amount of sampling data become available. As much descriptor information as possible with respect to climatic, soil, and other variables was collected for 1 square-mile vulnerable sections in California. Multivariate clustering techniques were then used to determine whether the descriptor information could be used to identify unique groups of vulnerable sections.

MATERIALS AND METHODS

Determination of Vulnerable Sections

A vulnerable section was defined as a 1 square-mile area of land where pesticide residues had been found in ground water due to agricultural use. By definition, all sections designated as Pesticide Management Zones (PMZs) in California were included, but other sections not regulated as PMZs were also included. Sections with bentazon and aldicarb detections were not designated as PMZs because their regulations apply statewide. Also, sections with detections of active ingredients that are no longer registered in California were not designated as PMZs (Maes, et. al., 1991). 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 a sampling conducted by the California Department of Health Services (Brown et al., 1986). Detections could have resulted from movement of contaminated ground water between sections during the time span between cessation of use and sampling of well water. This problem may be amplified for DBCP because of a long half-life in ground water, estimated at greater than 100 years (Burlinson et. al., 1982), and because large quantities were applied to soil.

Data for pesticide detections in well water, excluding DBCP detections, were obtained from the Well Inventory Data Base maintained since 1985 by the Department of Pesticide Regulation (DPR) (Cardozo et al.,1985). The Pesticide Prevention Contamination Act (Connelly, 1986) requires the DPR to determine whether or not reported detections are due to agricultural use. Therefore, detections determined to be due to agricultural use were used as indicators of areas that are vulnerable to contamination of ground water as a result of nonpoint agricultural use of pesticides. A total of 258 sections were identified as KV sections.

Data Sources

Climatic data for temperature and precipitation were obtained from a weather station database maintained by the California Department of Water Resources (CDWR). Data were obtained from 127 weather stations. Mean values for cumulative and monthly rainfall and for mean yearly and monthly temperature were derived from daily values averaged over 30 years at each station for 1961-1990 (Table 1). The weather station closest to the center of each KV section was determined from latitude-longitude coordinates.

Data for physical and chemical properties of soil were obtained at the level of soil mapping unit as delineated in soil survey maps for individual counties in California. The type of mapping unit used in this study was primarily surface texture phases of consociations of soil series (Soil Conservation Service, 1983). Two data sets were required. One data set identified the occurrence of soil mapping units in KV sections (personnel communication, Bob Teso, DPR, University of Riverside, Riverside, Ca). This data set was used to extract information from a second data, the Map Unit Interpretations Record (MUIR) data base provided by the Soil Conservation Service (SCS), USDA. The MUIR data base contains chemical, textural, and observational data by soil layer to the 5 foot depth for each soil mapping unit. Variables for soil texture in the MUIR database were presented in descriptive terms such as 'sandy loam'. These descriptions were transformed to a numeric scale by assigning values for sand and clay determined from the centroid of corresponding textural classes in the Soil Triangle (Soil Conservation Service, 1975) (Table 2). Other categorical variables whose categories were ordinal were transformed to a numeric scale. High and low values were reported for numeric variables so mid-points were calculated. The amount of data present for each variable varied between soil layers. Data for certain variables were partitioned to represent surface and subsurface conditions. The variable representative of the surface soil was derived by averaging data over the first soil layer for all soil mapping

Table 1. Description of climatic and soil variables.

Acronym	Description
Climate variables[a]	
MIYR	Yearly minimum temperature
MXYR	Yearly maximum temperature
MNYR	Average Yearly temperature
M1T-M12T	Mean monthly January-December temperature
PYR	Mean annual precipitation
M1P-M12P	Mean monthly January-December temperature
Soil variables[b]	
Textsand, Textclay	Derived percent sand and clay for Overall USDA Textural class of the soil
Txt1sand, Txt1clay	Derived percent sand and clay for USDA textural class for surface soil
Subtxsnd, Subtxcly	Derived percent sand and clay for USDA textural class for subsurface soil
Lay1clay, Subclay	Measured % clay content of surface and subsurface soil reported in the MUIR data base.
Lay1no4[c], Subno4[c]	Percent by weight of soil material smaller than 76 mm in diameter that passes a no. 4 (5 mm) sieve
Lay1no10[c], Subno10[c]	Percent by weight of soil material smaller than 76 mm in diameter that passes a no. 10 (2 mm) sieve
Lay1n200[c], Subn200[c]	Percent by weight of soil material smaller than 76 mm in diameter that passes a no. 200 (75 µm) sieve
Textind	Indicator for cobbles or stoniness in overall USDA textural class
Text1ind, Subtxind	Cobble or stoniness indicator for surface and subsurface soil
Lay1shsw, Subshsw	Shrink-swell of the surface and subsurface soil
Lay1perm, Subperm	Permeability of the surface and subsurface soil
Lay1deph, Subdeph	Average depth of the derived surface and subsurface soil layers
Lay1awc[c], Subawc[c]	Average water holding capacity of surface and subsurface soil
Lay1om[c]	Average percentage of organic matter in the surface
Hyd	Hydrologic group
Pan[c]	Indicator for hard pan
Drain[c]	Drainage group
Wattab[c]	Indicator for presence of a water table above 1.5 m
Flood[c]	Indicator for presence of annual flooding
Slope[c]	Surface slope of soil
DGW[a,c]	Depth from surface to ground water

a Data obtained from California Department of Water Resources.
b Data obtained from Soil Conservation Service.
c Variables used in the classification algorithm.

Table 2. Scale transformations used for soil variables.

Variable	Initial Scale	Transformed Scale
Texture	Sand	92% sand, 4% clay
	Loamy Sand	83% sand, 6% clay
	Sandy Loam	65% sand, 11% clay
	Loam	42% sand, 20% clay
	Silt Loam	20% sand, 15% clay
	Silt	8% sand, 6% clay
	Clay Loam	33% sand 34% clay
	Sandy Clay Loam	59% sand, 28% clay
	Silty Clay Loam	10% sand, 33% clay
	Sandy Clay	52% sand, 40% clay
	Silty Clay	7% sand, 46% clay
	Clay	20% sand, 60% clay
Water Table	No indication	0
	APPAR or PERCH	1
Annual Flooding	NONE	0
	RARE	1
	COMM or FREQ or OCCAS	2
Drainage Class	VP	0
	P	1
	SP	2
	MW	3
	W,MW	3.5
	W	4
	W,SE	4.5
	SE	5
	E	6
Hard Pan	No indication	0
	THICK or THIN	1
Shrink-Swell	LOW	0
	MODERATE	1
	HIGH	2
Hydrologic Group	A	0
	B	1
	C	2
	D	3

units within a section. The variable representative of the subsurface soil was derived by averaging data for all soil layers below the first layer within a mapping unit and then averaging across all mapping units within a section. Missing data for Del Norte, Humbolt, Kern and Tulare counties were obtained manually from published soil surveys or through personal contact with local SCS personnel. Soil data could not be obtained for KV sections in Los Angeles, Orange, and San Bernardino counties. This reduced the number of KV sections used in the statistical analysis from 258 to 180.

One other variable, depth to ground water, was obtained from a 1985 CDWR report that contained information for specific wells with PLSS Township-Range identifications. Since only a portion of vulnerable sections contained data, a gridding procedure, available in the SAS® statistical package was used to produce estimated values (SAS Inc, 1988). Del Norte, Humbolt, and Santa Clara Counties lacked enough information to conduct the gridding. Values for vulnerable sections in these areas were estimated from well log information. Depth to ground water could not be determined for 9 other KV sections. In the discussion that follows, depth to ground water will be grouped with soil data.

Each data set was initially processed using the ORACLE® database management system on a SUN® computer. The processed data were output to a single file in American Standard Code For Information Interchange (ASCII) format with each record representing a vulnerable section and containing all climate and soil data for that section. Twenty-eight climatic and thirty-three soil variables were identified (Table 1). The ASCII data file was analyzed with SAS® software on a DOS based personal computer (SAS Institute, 1988).

Cluster Analysis

Initially, the plan was to conduct a cluster analysis using all climate and soil variables. Climate variables, however, dominated results of the first analysis. This was caused by a difference in the variance structure between climate and soil variables for KV sections. Since weather stations were less numerous than KV sections, identical rainfall and temperature data were assigned to some KV sections. When means were obtained for each county, the variance for climate variables was zero. In order to retain climate information for KV sections, a two-stage process was developed where in the first, cluster analysis was conducted on climate variables from 32 weather stations nearest KV vulnerable sections. In the second stage, cluster analysis was conducted on soil variables from KV sections within climate clusters.

When the number of variables is large, one common clustering procedure is to first conduct a Principal Components Analysis (PCA) analysis on all variables to determine if a subset of principal components (PCs) could be used to describe the raw data set. Clustering procedures are then conducted on the reduced number of principal components (Gnanadesikan and Kettenring, 1989). This procedure has two disadvantages. First, description of the clusters could be unclear because assignment of meaning to the principal components could be difficult. Second, use of principal components could produce indistinct clustering results and obscure the actual number of clusters that exist in a data set (Fowlkes, et. al., 1988). The latter was observed with the soils data.

An alternative procedure was developed 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. 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 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). A correlation coefficient value < 0.75 was selected as the cut-off point for inclusion. This process was repeated until there was no clear clustering from the higher-order combinations of variables.

Three statistical measures were used to determine the number of clusters; the Cubic Clustering Criterion (CCC), the Pseudo-F and Pseudo-t statistics (SAS Institute Inc., 1983; SAS Institute Inc., 1988). Three clustering methods were used: Ward, Average linkage, and Centroid. In the Ward method, distance between two clusters is computed as the Analysis of Variance sum of squares between the two clusters added up over all the variables. 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 criteria and between methods.

Classification of Candidate Vulnerable Sections

Vulnerability classification was based on measuring the similarity of soil data from candidate sections to profiles developed from the clustering analysis. Climate data would be used as a screen to determine the appropriate soil profile test. Soil profiles were developed by conducting a PCA analysis on the standardized soil variables within identified clusters, and then computing the mean and standard deviation of each principal component (PC) score. Corresponding PCs for each cluster would be calculated for soil data from candidate sections. Inclusion of a candidate section into one of the vulnerable clusters would occur only if every PC score from the candidate section fell within a specified distance of the corresponding cluster's PC mean. The distance for each PC was determined as a constant 'K' multiplied by the cluster standard deviation of the PC. The value of K was chosen by examining the proportion of correct and incorrect classifications of KV sections as a function of K.

RESULTS

Climate Variables

Prior to clustering, correlation analysis was conducted on climatic variables from 32 weather stations nearest KV sections. In general, temperature variables were uncorrelated with precipitation variables (Table 3). Minimum yearly temperature was highly correlated with September through May mean monthly temperatures and less correlated with June, July, and August values. In contrast, maximum yearly temperature was highly correlated with March through October mean monthly temperatures and less correlated with November through February values. Total annual precipitation was highly correlated with September through May monthly

Table 3. Correlation matrix for climate variables. Correlation coefficients of 0.75 or greater are underlined to illustrate trends in the data. Acronyms are defined in Table 2.

Pearson Correlation Coefficients/N=32

	Temperature Variables															Precipitation Variables												
	MIYR	MXYR	MNYR	M1T	M2T	M3T	M4T	M5T	M6T	M7T	M8T	M9T	M10T	M11T	M12T	PYR	M1P	M2P	M3P	M4P	M5P	M6P	M7P	M8P	M9P	M10P	M11P	M12P
MIYR	1.00	0.68	0.90	0.88	0.95	0.96	0.89	0.76	0.64	0.56	0.64	0.78	0.92	0.94	0.85	-0.13	-0.03	-0.01	-0.03	-0.07	-0.35	-0.71	-0.74	-0.27	-0.22	-0.23	-0.13	-0.14
MXYR		1.00	0.93	0.49	0.66	0.81	0.92	0.95	0.94	0.93	0.96	0.97	0.90	0.64	0.45	-0.60	-0.48	-0.50	-0.57	-0.59	-0.72	-0.78	-0.71	-0.33	-0.62	-0.64	-0.63	-0.59
MNYR			1.00	0.73	0.86	0.96	0.99	0.94	0.87	0.82	0.88	0.96	0.99	0.85	0.69	-0.42	-0.30	-0.30	-0.35	-0.38	-0.60	-0.81	-0.79	-0.33	-0.47	-0.49	-0.43	-0.41
M1T				1.00	0.97	0.87	0.70	0.47	0.31	0.23	0.34	0.53	0.78	0.98	1.00	0.02	0.13	0.14	0.16	0.08	-0.24	-0.63	-0.60	-0.17	-0.05	-0.14	0.00	0.01
M2T					1.00	0.96	0.85	0.67	0.52	0.43	0.53	0.70	0.90	1.00	0.95	-0.11	0.02	0.02	0.01	-0.05	-0.39	-0.75	-0.72	-0.28	-0.21	-0.25	-0.13	-0.11
M3T						1.00	0.96	0.83	0.72	0.63	0.71	0.85	0.97	0.95	0.84	-0.27	-0.15	-0.14	-0.18	-0.21	-0.52	-0.83	-0.79	-0.34	-0.37	-0.38	-0.29	-0.27
M4T							1.00	0.95	0.88	0.82	0.87	0.95	0.98	0.83	0.66	-0.42	-0.30	-0.31	-0.36	-0.37	-0.61	-0.84	-0.80	-0.36	-0.50	-0.49	-0.44	-0.41
M5T								1.00	0.98	0.94	0.96	0.98	0.91	0.63	0.42	-0.52	-0.42	-0.42	-0.50	-0.49	-0.64	-0.76	-0.74	-0.35	-0.57	-0.54	-0.53	-0.50
M6T									1.00	0.99	0.98	0.97	0.83	0.49	0.26	-0.57	-0.48	-0.49	-0.58	-0.56	-0.64	-0.67	-0.67	-0.32	-0.59	-0.55	-0.57	-0.55
M7T										1.00	0.99	0.94	0.77	0.41	0.18	-0.63	-0.56	-0.57	-0.64	-0.63	-0.64	-0.61	-0.60	-0.28	-0.60	-0.60	-0.63	-0.62
M8T											1.00	0.97	0.83	0.51	0.30	-0.62	-0.53	-0.54	-0.61	-0.61	-0.65	-0.67	-0.65	-0.29	-0.59	-0.61	-0.62	-0.61
M9T												1.00	0.93	0.68	0.48	-0.54	-0.44	-0.44	-0.51	-0.53	-0.66	-0.76	-0.74	-0.33	-0.57	-0.58	-0.55	-0.53
M10T													1.00	0.89	0.75	-0.39	-0.26	-0.26	-0.31	-0.35	-0.60	-0.83	-0.80	-0.36	-0.46	-0.48	-0.41	-0.38
M11T														1.00	0.97	-0.11	0.02	0.02	0.01	-0.06	-0.38	-0.73	-0.71	-0.26	-0.20	-0.26	-0.14	-0.12
M12T															1.00	0.06	0.16	0.17	0.19	0.10	-0.21	-0.59	-0.57	-0.15	-0.02	-0.11	0.03	0.04
PYR																1.00	0.97	0.98	0.98	0.97	0.87	0.59	0.47	0.44	0.88	0.97	0.99	0.99
M1P																	1.00	0.99	0.95	0.93	0.73	0.43	0.32	0.29	0.75	0.89	0.94	0.98
M2P																		1.00	0.97	0.96	0.77	0.45	0.33	0.30	0.80	0.91	0.96	0.98
M3P																			1.00	0.98	0.85	0.49	0.39	0.39	0.88	0.93	0.97	0.96
M4P																				1.00	0.84	0.50	0.37	0.32	0.84	0.93	0.96	0.96
M5P																					1.00	0.83	0.71	0.64	0.96	0.93	0.90	0.83
M6P																						1.00	0.90	0.69	0.74	0.70	0.61	0.55
M7P																							1.00	0.76	0.69	0.60	0.49	0.45
M8P																								1.00	0.72	0.56	0.46	0.39
M9P																									1.00	0.92	0.90	0.84
M10P																										1.00	0.98	0.96
M11P																											1.00	0.98
M12P																												1.00

average precipitation and less correlated with June, July, and August values. The forward clustering technique identified 5 distinct clusters formed from 3 variables. The clustering variables, given in order of selection, were average January temperature, average March precipitation, and average July precipitation. Means for each variable in each cluster are given for the solution derived from the Ward method (Table 4). Clusters 3 and 5 had high precipitation values: cluster 5 had the highest March precipitation and cluster 3 the highest July precipitation. Clusters 1, 2, and 4 had low precipitation vales, differing mainly in January temperatures: cluster 2 had the highest and cluster 4 the lowest January temperatures. The following geographic patterns were observed when weather station membership in each cluster was identified by county location of the weather station (Table 5). Cluster 1 was dominated by counties in the the Central Valley. Counties in cluster 2 were located in the central and south coasts and in inland portions of southern California. Counties in clusters 3 and 5 were Humbolt and Del Norte, northern coastal counties. Siskiyou comprised cluster 4, reflecting the weather of a higher mountainous locale.

Soil Variables

Theoretically, clustering of soil variables would have occurred within each of the climate clusters to identify unique soil clusters within climate clusters. There were insufficient numbers of sections in most of the climate clusters to perform this analysis. However, the results of the climate clustering were highly indicative that KV sections in clusters 1, 2, and 4 could be grouped because they represented a low rainfall condition when compared to much higher rainfall values for those in clusters 3 and 5. Thus, the eleven sections in Del Norte and Humbolt counties were excluded from the soil clustering analysis. An additional 9 sections were excluded

Table 4. Means by cluster for weather variables produced by the 5 cluster solution for the Ward clustering method.

Cluster	Weather Stations	Nearest KV Sect	January Temperature	Precipitation March	Precipitation July
	#	#	°F	inches	inches
1	20	153	45	1.8	0.03
2	7	16	52	2.0	0.04
3	2	2	47	4.5	0.71
4	2	2	29	1.0	0.29
5	1	7	47	8.6	0.31

Table 5. Cluster association given by county location for 32 weather stations nearest KV sections.

Cluster	County Location of Weather Station
1	Contra Costa, Colusa, Fresno, Glenn, Kern, Merced, Sacramento, San Joaquin, Stanislaus, Tehama, Tulare, Yolo, Yuba
2	Santa Cruz, Orange, Riverside, Santa Clara, San Diego
3	Humbolt
4	Siskiyou
5	Del Norte

because of a lack of depth to ground water data. A total of 160 KV sections located in dry weather clusters were used in the soil clustering analysis.

Correlation analysis was first conducted on the 33 soil variables (Table 6). One group consisted of 15 highly correlated variables which was comprised of 10 variables that indicated texture in terms of sand and clay content of either the surface or subsurface soil, 4 variables that measured the permeability and shrink-swell potential of the surface and subsurface soil, and a variable that indicated the hydrologic category of the soil. A second group of seven correlated variables consisted of indicators of cobbly or stony soil and measures of the percentage by weight of soil particles passing through coarse sieve sizes Nos. 4 and 10. The 11 remaining variables were uncorrelated.

The best clustering variable in the first step of the forward selection technique was a texture variable that measured the percent by weight of soil particles that pass through a No. 200 soil sieve. Soil texture is reflected by this variable in the following way: the lower the number of soil particles passing through the No. 200 sieve, the greater the sand content of the soil and conversely, the greater the number, the greater the clay content of the soil. Two clusters were indicated with this single variable. The best combinations of variables that indicated clustering in subsequent steps are given in Table 7. The final solution occurred with a combination of four variables: 1) the texture variable measuring soil particles passing a No. 200 soil sieve; 2) a variable that indicated presence of a water table above 5 feet some time during the year;

Table 6. Correlation matrix for soil variables. Correlation coefficients of 0.75 or greater are underlined to illustrate trends in the data. Acronyms are defined in Table 2.

Pearson Correlation Coefficients/N=160

Variables Correlated with Soil Texture

	TEXTSAND	TEXTCLAY	TXT1SND	TXT1CLY	LAY1SHSW	HYD	LAY1CLAY	LAY1PERM	LAY1N200	SUBTXSND	SUBTXCLY	SUBSHSW	SUBCLAY	SUBPERM	SUBN200
TEXTSAND	1.00	-0.87	0.98	-0.87	-0.82	-0.80	-0.89	0.89	-0.95	0.90	-0.85	-0.86	-0.87	0.79	-0.88
TEXTCLAY		1.00	-0.87	0.99	0.97	0.85	0.98	-0.78	0.91	-0.88	0.92	0.92	0.92	-0.73	0.89
TXT1SND			1.00	-0.88	-0.84	-0.80	-0.91	0.89	-0.96	0.92	-0.86	-0.87	-0.89	0.79	-0.90
TXT1CLY				1.00	0.97	0.84	0.99	-0.78	0.92	-0.88	0.92	0.92	0.92	-0.72	0.90
LAY1SHSW					1.00	0.76	0.97	-0.68	0.90	-0.86	0.89	0.91	0.89	-0.63	0.89
HYD						1.00	0.85	-0.84	0.83	-0.84	0.90	0.88	0.89	-0.87	0.82
LAY1CLAY							1.00	-0.82	0.94	-0.92	0.94	0.94	0.95	-0.76	0.93
LAY1PERM								1.00	-0.85	0.86	-0.82	-0.79	-0.84	0.89	-0.80
LAY1N200									1.00	-0.96	0.92	0.93	0.94	-0.79	0.96
SUBTXSND										1.00	-0.96	-0.95	-0.97	0.86	-0.98
SUBTXCLY											1.00	0.99	0.99	-0.82	0.95
SUBSHSW												1.00	0.99	-0.79	0.95
SUBCLAY													1.00	-0.84	0.97
SUBPERM														1.00	-0.81
SUBN200															1.00
TEXTIND															
TXT1IND															
LAY1NO4															
LAY1NO10															
SUBTXIND															
SUBNO4															
SUBNO10															
LAY1DEPH															
PAN															
SUBDEPH															
DRAIN															
WATTAB															
FLOOD															
SLOPE															
LAY1OM															
LAY1AWC															
SUBAWC															

Table 6. Continued.

Pearson Correlation Coefficients/N=160

	Variables Correlated with Coarse Soil Variables							Uncorrelated Variables										
	TEXTIND	TXT1IND	LAY1NO4	LAY1NO10	SUBTXIND	SUBNO4	SUBNO10	LAY1DEPH	PAN	SUBDEPH	DRAIN	WATTAB	FLOOD	SLOPE	LAY1OM	LAY1AWC	SUBAWC	DGW
TEXTSAND	-0.04	-0.03	-0.26	-0.41	-0.02	-0.22	-0.34	-0.45	-0.05	-0.28	0.64	-0.57	-0.11	0.08	-0.69	-0.69	-0.58	0.45
TEXTCLAY	0.01	-0.01	0.28	0.43	-0.07	0.29	0.41	0.61	-0.03	0.35	-0.62	0.58	0.10	0.00	0.65	0.53	0.55	-0.41
TXT1SND	-0.05	-0.03	-0.27	-0.42	-0.02	-0.24	-0.36	-0.47	-0.01	-0.29	0.69	-0.63	-0.16	0.09	-0.72	-0.69	-0.61	0.44
TXT1CLY	0.02	0.00	0.27	0.42	-0.06	0.29	0.42	0.62	-0.05	0.36	-0.61	0.58	0.07	0.00	0.64	0.53	0.54	-0.41
LAY1SHSW	-0.05	-0.07	0.34	0.49	-0.12	0.36	0.48	0.59	-0.15	0.36	-0.59	0.59	0.12	-0.04	0.64	0.48	0.55	-0.39
HYD	0.04	0.04	0.22	0.34	-0.04	0.23	0.33	0.58	0.23	0.28	-0.63	0.44	-0.08	0.10	0.49	0.57	0.44	-0.47
LAY1CLAY	0.03	0.01	0.27	0.43	-0.05	0.28	0.41	0.61	-0.07	0.37	-0.65	0.61	0.12	-0.03	0.66	0.55	0.58	-0.38
LAY1PERM	-0.14	-0.13	-0.08	-0.20	-0.13	-0.04	-0.14	-0.51	-0.14	-0.29	0.61	-0.44	-0.06	0.00	-0.55	-0.70	-0.60	0.29
LAY1N200	-0.02	-0.03	0.35	0.51	-0.05	0.31	0.43	0.55	-0.05	0.33	-0.67	0.62	0.11	-0.11	0.70	0.68	0.64	-0.48
SUBTXSND	-0.02	-0.01	-0.28	-0.44	0.05	-0.26	-0.39	-0.54	0.08	-0.28	0.68	-0.59	-0.13	0.09	-0.64	-0.65	-0.66	0.43
SUBTXCLY	0.04	0.03	0.25	0.41	-0.07	0.27	0.40	0.59	-0.04	0.31	-0.64	0.52	0.05	-0.01	0.58	0.56	0.55	-0.43
SUBSHSW	-0.01	-0.02	0.30	0.46	-0.13	0.33	0.46	0.58	-0.05	0.31	-0.63	0.53	0.08	-0.02	0.60	0.56	0.54	-0.43
SUBCLAY	0.03	0.02	0.27	0.43	-0.08	0.29	0.43	0.59	-0.05	0.29	-0.67	0.55	0.08	-0.03	0.62	0.58	0.58	-0.42
SUBPERM	-0.04	-0.03	-0.15	-0.26	0.00	-0.15	-0.24	-0.51	-0.19	-0.26	0.62	-0.39	-0.01	-0.03	-0.49	-0.63	-0.62	0.29
SUBN200	-0.07	-0.09	0.38	0.54	-0.15	0.38	0.51	0.56	-0.13	0.29	-0.67	0.60	0.12	-0.09	0.65	0.61	0.67	-0.43
TEXTIND	1.00	0.99	-0.86	-0.76	0.86	-0.80	-0.72	0.01	-0.10	-0.12	0.11	-0.04	-0.10	0.07	0.00	0.00	-0.07	-0.03
TXT1IND		1.00	-0.85	-0.76	0.85	-0.80	-0.72	0.00	-0.09	-0.13	0.12	-0.05	-0.10	0.07	0.00	-0.01	-0.09	-0.02
LAY1NO4			1.00	0.95	-0.78	0.91	0.87	0.13	0.09	0.10	-0.34	0.30	0.08	-0.09	0.29	0.25	0.23	-0.31
LAY1NO10				1.00	-0.73	0.89	0.91	0.21	0.01	0.14	-0.46	0.42	0.15	-0.18	0.42	0.32	0.32	-0.33
SUBTXIND					1.00	-0.89	-0.84	-0.03	-0.02	0.02	0.15	-0.06	-0.14	0.01	-0.09	0.02	-0.04	-0.02
SUBNO4						1.00	0.98	0.14	0.01	0.01	-0.33	0.26	0.06	0.00	0.29	0.13	0.19	-0.22
SUBNO10							1.00	0.22	-0.06	0.04	-0.41	0.35	0.09	-0.02	0.38	0.18	0.27	-0.25
LAY1DEPH								1.00	-0.12	0.66	-0.29	0.24	-0.13	0.04	0.33	0.33	0.38	-0.24
PAN									1.00	-0.11	-0.09	-0.13	-0.15	0.10	0.00	0.17	-0.07	-0.13
SUBDEPH										1.00	-0.18	0.17	0.06	-0.24	0.16	0.12	0.13	-0.04
DRAIN											1.00	-0.85	-0.43	0.19	-0.67	-0.36	-0.35	0.33
WATTAB												1.00	0.57	-0.28	0.65	0.29	0.33	-0.34
FLOOD													1.00	-0.36	0.35	0.07	0.05	0.21
SLOPE														1.00	-0.08	-0.07	0.05	-0.04
LAY1OM															1.00	0.58	0.52	-0.44
LAY1AWC																1.00	0.66	-0.46
SUBAWC																	1.00	-0.23
DGW																		1.00

Table 7. Comparison between clustering methods and between criteria for the number of clusters found from stepwise addition of soil variables. Acronyms are defined in Table 2.

Variables and Clustering Method	Number of Clusters According to these Criteria		
	CCC	Pseudo-F	Pseudo-t
Step 1: Lay1n200[a]			
Ward	2	2	5
Average	2	2	2
Centroid	2	2	2
Step 2: Lay1n200 and Wattab			
Ward	3	3	3
Average	3	3	3
Centroid	5	5	4
Step 3: Lay1n200, Wattab, and Slope			
Ward	4	4	4
Average	5	5	5
Centroid	5	5	5
Step 4: Lay1n200, Wattab, Slope, and Lay1no4			
Ward	6	5	5
Average	7	7	7
Centroid	7	7	7

3) a variable that indicated the average slope of the section; and 4) a variable that measured the number of soil particles that pass through a No. 4 soil sieve (the lower the number, the more volume of soil taken up by large soil constituents). Although the number of clusters differed between the Ward and the other 2 clustering methods at the four variable solution, the variables selected by the methods were identical. The Average and Centroid methods indicated a 7 cluster solution but 2 extra clusters, enclosed in the boxes in Figure 1, were produced from an early split of the same parent clusters identified in the Ward procedure.

The 5 cluster solution from the Ward method was determined as the final solution. Each cluster from this solution had a unique combination of variables as indicated by the means for variables in each cluster (Table 8). Soils in clusters 1 and 3 were clayey, as indicated by the higher % values for the Lay1n200 variable and had shallow slopes. Cluster 3 was split from cluster 1 because those sections also had a high incidence of soils with a water table above 5 feet. In contrast, soil in cluster 2 was sandy with shallow slope and with practically no presence of a shallow water table. Clusters 4 and 5 were intermediate in terms of surface soil texture but each was unique in that sections in cluster 4 had greater values for slope and those in cluster 5 had a greater incidence of large soil particles such as cobbles or stones as indicated by the lower % values for the Lay1no4 variable.

Assessment of the clustering results was conducted by mapping the location of sections as identified by cluster association. There was good geographic

Figure 1. Heirarachical clustering results for the Ward and Average methods using 4 soil variables. Underlined numbers represent the final cluster solution for each method and numbers inside boxes for Average method are splits of clusters 11 and 15 in Ward method.

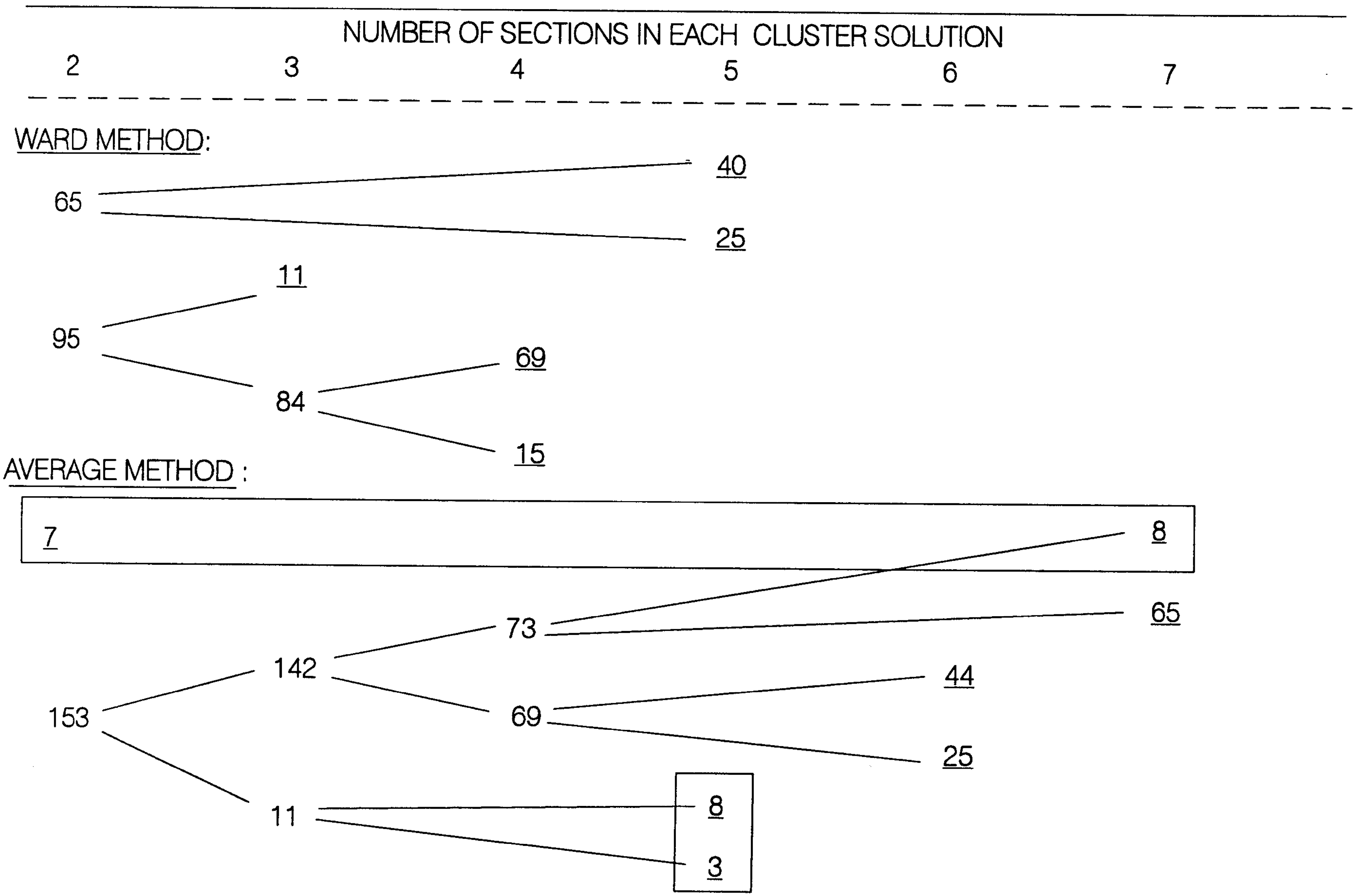


Table 8. Means by cluster for soil variables produced by the 5 cluster solution for the Ward clustering method.

Cluster	N	Lay1n200	Wattab[a]	Slope	Lay1no4
		---%----		---------%-----------	
1	40	79	0.22	1.4	99
2	69	40	0.04	1.6	96
3	25	81	0.76	0.8	98
4	11	57	0	12.7	96
5	15	56	0.15	2.6	86

a Scale from 0-1 with a 0 value representing no soils in a section with a shallow water table above 5 feet and a value of 1 representing all soils in a section with a shallow water table.

separation between clusters. Sandy sections in cluster 2 were predominately located in the southern portion of the Central Valley and in the Southern Desert areas whereas clayey sections in clusters 1 and 3 were predominately located in the northern portion of the Central Valley (Figure 2). Within the clayey clusters, those with a greater incidence of shallow water table were located in a band sandwiched between groups of sections in cluster 1 (Figure 3). Sandy sections of cluster 2 in the southern Central Valley were located along the valley floor with some sections in cluster 4 located along the foothills (Figure 4). Thus, the clustering appeared effective in providing a regional description of the location of vulnerable sections. If pathways of contamination are related to variables associated with each cluster, then it may be possible to devise and specify cluster-based management strategies. This approach could facilitate management decisions on a regional basis.

Procedure for Identifying Vulnerable Sections

A two-stage procedure for identifying candidate sections as vulnerable was developed. The first stage would be a climate screen to determine if the candidate section's rainfall was either high or low. If the candidate section had high rainfall, then it would be subject to a soil profile test derived from soil data for KV sections in Humbolt and Del Norte counties. If the candidate section had low rainfall, then it would be subjected to further classification based on the soil profiles developed from KV sections in each of the 5 low-rainfall soil clusters.

Figure 2. Spatial location of sections in clusters 1 and 3 with predominantly clayey soil contrasted to locations of sections in cluster 2 with predominantly sandy soil.

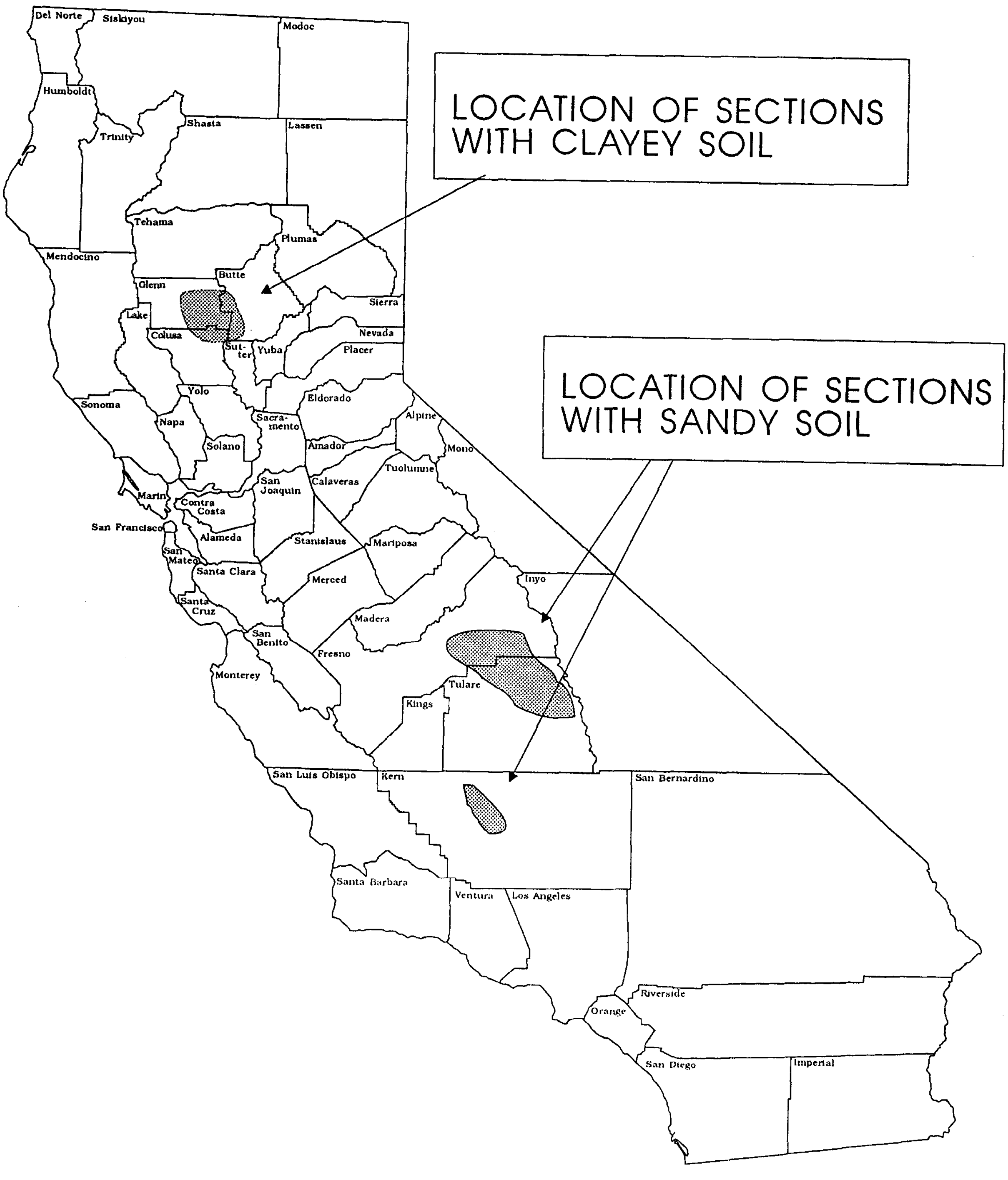


Figure 3. Cluster membership for known vulnerable sections in the northern Central Valley.

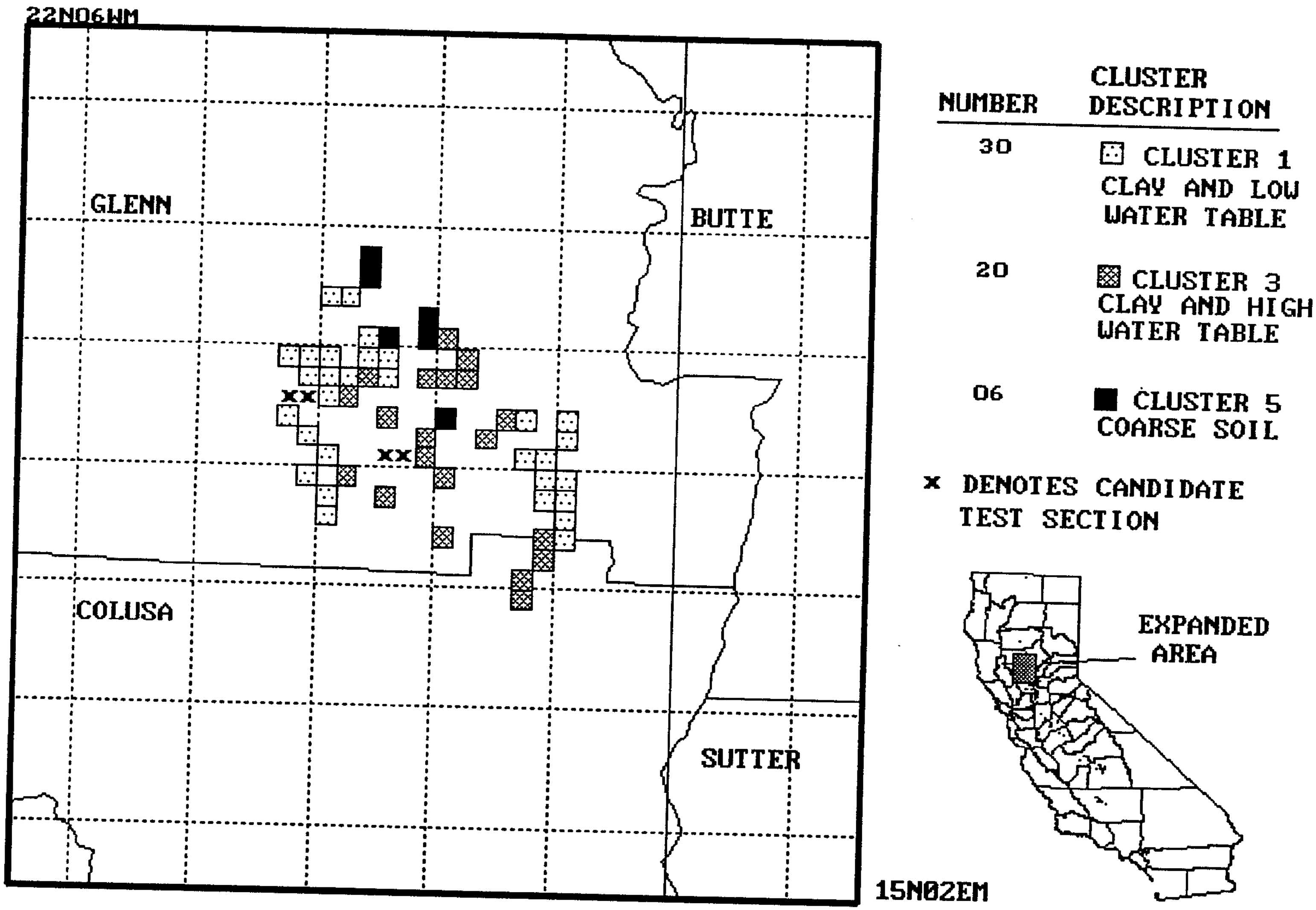
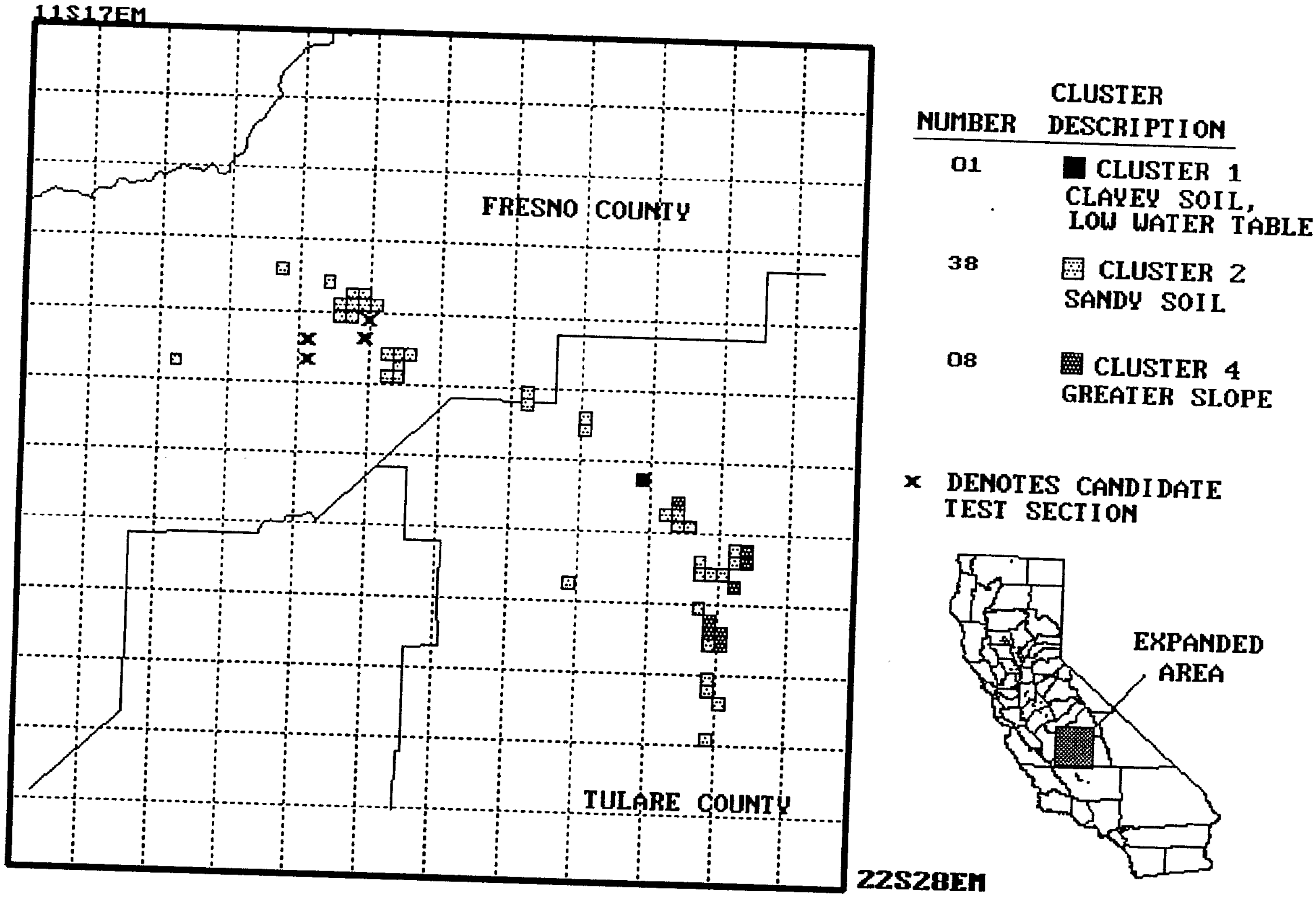
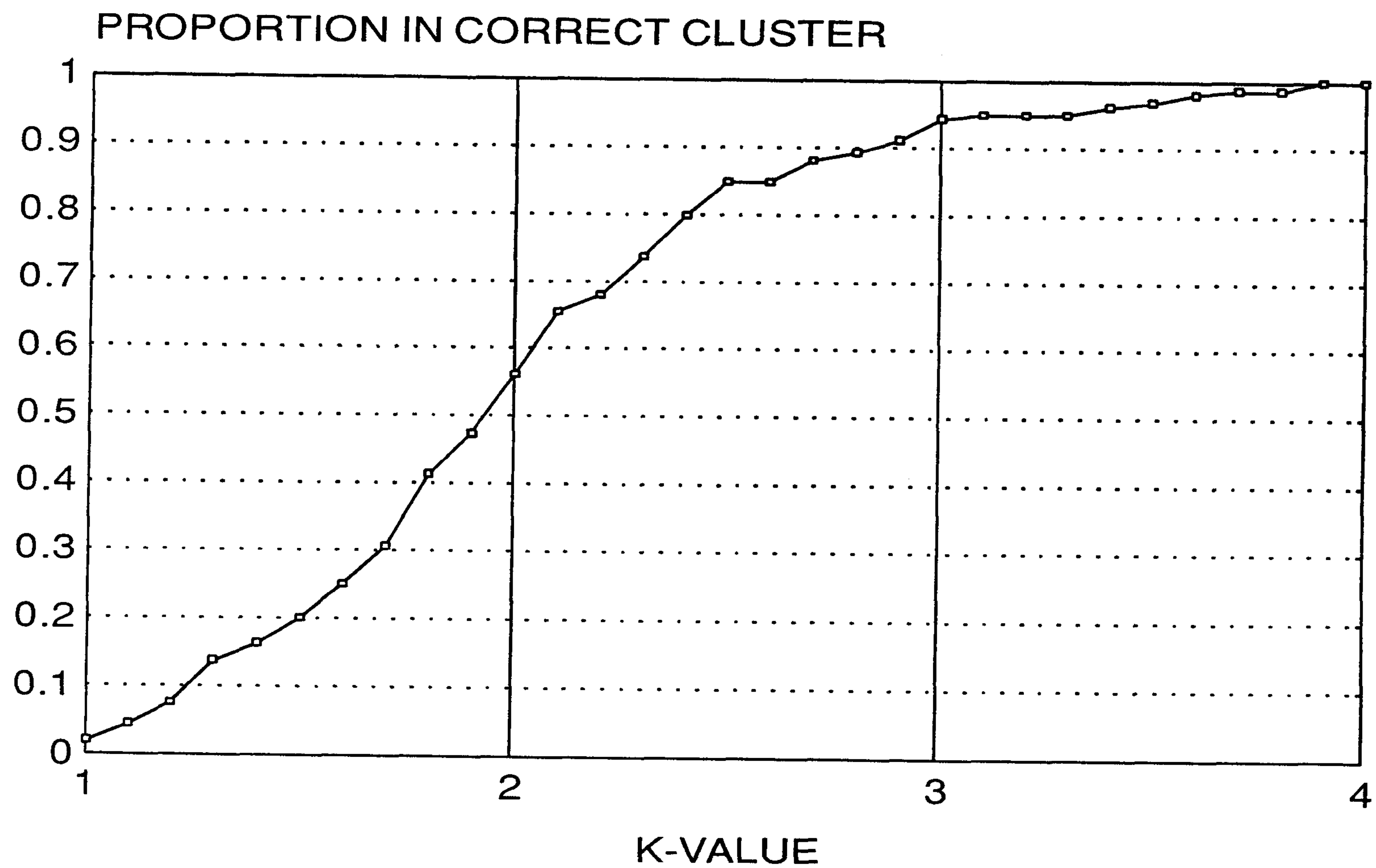


Figure 4. Cluster membership for known vulnerable sections in the southern Central Valley in Fresno and Tulare counties.



Soil profiles were developed using 15 of the 33 soil variables. Redundant variables were excluded from the algorithm. For example, the number of soil particles passing sieve No. 200 was highly correlated with all derived texture variables so the derived texture variables were omitted. Classification into surface and subsurface layers was retained because this could be an informative division in future investigations. The variables denoted with the superscript 'c' in Table 1 were used to develop the classification algorithm. For a candidate section, data for the 15 soil variables first would be standardized to the mean and standard deviation of corresponding variables in each of the KV soil clusters. PC scores would be calculated for the candidate section based on the 15 standardized variables and the values compared to a specified range for each corresponding PC in a soil cluster. Inclusion into a cluster would occur only if <u>all</u> PC scores were within 3 standard deviations of the mean (zero). The multiplier 3 was determined from a plot of the number of KV sections correctly classified as a function of the value of K. When the value was 3, 95% of the KV sections were correctly classified into their respective soil clusters (Figure 5). Although a larger value of K would achieve 100% correct classifications, it could also result in the classification of more dissimilar sections as vulnerable. A candidate section would be classified as vulnerable if it could be considered a member of one of the soil clusters, otherwise it would be considered as not classifiable. There is no implication that sections not classified are invulnerable.

Figure 5. Proportion of sections classified into the correct cluster as a function of K, a multiplier of the standard deviation of PC scores.



An example of the classification procedure is given for eight candidate sections, 4 in Glenn county and 4 in Fresno county. The sections were chosen from areas near 3 of the low-rainfall KV clusters (Figures 3 and 4). Since these sections are near low rainfall weather stations, their soil data were compared to the 5 clusters identified from the cluster analysis of the low rainfall KV sections. The occurrence of soil mapping units in each of the candidate sections was manually determined from SCS soil maps in published soil surveys for Glenn and Fresno counties. Data for the 15 soil variables for each of the soil mapping units were extracted from the MUIR database and average values calculated for each section. The average sectional values were standardized to the corresponding mean and standard deviation of each of the 5 soil clusters. Next, PC scores for the standardized values were calculated by multiplying the standardized values with the PC coefficients for each of the 5 clusters. The membership test was then conducted by determining if each sectional PC score was within 3 standard deviations of the mean of that cluster. Results in Table 9 are expressed in terms of the number of tests for that cluster where the PC score for the standardized section was outside the range. A value of zero indicates cluster membership. All 4 sections in Fresno county were classified as belonging to Cluster 2, the predominately sandy cluster. Two of the sections in Glenn county were geographically near and subsequently, classified into Cluster 3, clayey sections with high incidence of a water table above 5 feet. Two other section in Glenn county were geographically near Cluster 1, clayey sections with low incidence of a shallow water table, but only one of those sections was identified as a member of that cluster.

Table 9. Test of the classification of candidate sections for membership in one of the low rainfall soil clusters.

Section Location	Number of PC Tests Out of The Range for Cluster: 1	2	3	4	5
Fresno County					
15S21E01	6	0	12	2	6
15S21E06	6	0	12	6	7
15S21E07	6	0	10	6	6
15S21E12	5	0	10	3	6
Glenn County					
19N03W34	3	6	0	9	7
19N03W35	1	7	0	7	8
19N04W13	2	4	7	3	5
19N04W14	0	3	11	6	4

These results have two implications with respect to implementation of management strategies. First, the choice of a section as a basic geographic unit appears to give good results: averaging all mapping units within a section produced logical patterns with respect to geographic association of soil mapping units. Second, regional management strategies may be possible based on the clusters. For example, division of sections based on soil texture suggested that different management strategies may be required for clayey vs sandy soils: special properties of clay soil such as the appearance of cracks or a shallow water table could require a different set of management conditions than those generated for sandy soils. However, more information is needed on the processes important in pesticide movement in each cluster in order to provide a link between management practices and cluster identification.

In summary, the present study has endeavored to create profiles of groups of known contaminated sections in California with respect to a series of climatic and physical soil properties. The following question has been answered: what are <u>some</u> of the vulnerable sections in California like? The profile analysis of this study differed from a typical discriminant analysis in two ways. First, profiles were devised only for vulnerable sections: no non-vulnerable sections were studied or defined. We, therefore, have no way of evaluating the usefulness of the climatic and soil variables as discriminators for vulnerability. It is possible that variables were used which are not effective in discrimination. Second, profiles were created for five clusters comprising a total of only 160 sections. There may be other vulnerable cluster profiles with characteristics not included in our description of vulnerability. Therefore, new candidate sections which are not classified as similar to one of the known vulnerability clusters are not necessarily invulnerable. They retain a status of unknown vulnerability which could be changed when the clustering and classification procedures are updated to reflect new positive well sampling data.

SUMMARY

1. Clustering methods were successful in grouping vulnerable sections based on climate and soil variables. However, due to differences in the variance structure two separate procedures were used.
2. Clustering of data from weather stations resulted in 5 distinct groups that were related to geographic location of the weather station. With respect to pesticide movement to ground water, two of the clusters had high rainfall and contained 11 of the 180 vulnerable sections, 7 in Del Norte and 4 in Humbolt counties. The remaining KV sections were in the other 3 clusters that had low rainfall.
3. Clustering analyses were conducted on soil data from 160 KV sections that were members of the low rainfall clusters. Using a forward selection technique, four soil variables were identified that clustered 160 vulnerable sections into 5 groups. The variables were: 1) soil texture as measured by the percentage by weight of soil material smaller than 76 mm that passes a No. 200 (75 µm) soil sieve; 2) indication of the presence of a water table above 5 feet during the year; 3) the average slope of mapping units in a section; and 4) an indication of the presence of coarse soil particles such as cobbles or stones as measured by the percentage by weight of soil material smaller than 76 mm that passes a No. 4 (5 mm) soil sieve.

4. Due to differences noted in the variance of the climate and soils data sets, a two-stage approach was developed to identify candidate sections as vulnerable. A candidate section would be screened to determine whether if it had low or high rainfall. If the candidate section had low rainfall, then it would be subjected to a classification algorithm developed from the results of the clustering of soil variables for 160 vulnerable sections in low-rainfall areas. If the section passed the soil algorithm then it would be identified as a vulnerable section. If the section had high rainfall, then it would be considered vulnerable if data from soil variables passed an algorithm developed from properties of soils that occur in vulnerable sections in Del Norte and Humbolt counties.

References

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

Aller, L., T. Bennett, J.H. Lehr, and R.J. Petty. 1985. DRASTIC: A standardized system for Evaluating Ground Water Pollution Potential Using Hydrogeologic Settings. EPA/600/2-85/018, May 1985, Environmental Protection Agency, Washington D.C.

Balu, K., and R.T. Paulsen. 1991. Interpretation of atrazine in ground water data using a geographic information system. In D.L. Weigmann (ed.) Pesticides in the next decade: the challenges ahead. Proceed. of the Third National Research Conference on Pesticides, 8-9 Nov., 1990, Virginia Water Resources Research Center, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Cardozo, C.L., S. Nicosia, and J. Troiano. 1985. Agricultural pesticide residues in California well water: development and summary of a well inventory data base for non-point sources. California Department of food and Agriculture, Environmental Hazards Assessment Program, Sacramento, Ca.

Brown, M., C. Cardozo., S. Nicosia, J. Troiano, and S. Ali. 1986. Sampling for pesticide residues in California well water: 1986 well inventory data base. 1 Dec. 1986. First Annual Report to the Legislature, State Department of Health Services, and State Water Resources Control Board. Environmental

Hazards Assessment Program, California Department of Food and Agriculture, Sacramento, CA.

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

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. In Surveying theory and practice. Fifth edition. New York, N.Y.

EPA. 1992. ANOTHER LOOK: National pesticide survey phase II report. EPA 570/9-91-020 Jan., 1992. Office of Water, EPA, Washington D.C.

Fisher, C., and M.L. Reid. 1986. Ground water "Hot Spots:" toxic pollutant problem identification, correction, and prevention. California State Water Resources Control Board. Sacramento, Ca.

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

Gnanadesikan, R., and J.R. Kettenring. 1989. Discriminant analysis and clustering, Panel on discriminant analysis, classification and clustering. Statistical Sci., 4:34-69.

Holden, L.R., J.A. Graham, R.W. Whitmore, W.J. Alexander, 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.

Kissel, D.E., O.W. Bidwell, and J.F. Kientz. 1982. Leaching classes of Kansas soils. Kansas State University, Kansas Agricultural Experiment Station, Bulletin 641, Manhattan, KS.

Maes, C.M., M. Pepple, J. Troiano, D. Weaver, W. Kimaru, and SRCB staff. 1991. Sampling for pesticide residues in California well water 1991 update well inventory data base. Environmental Hazards Assessment Program, California Environmental Protection Agency. EH 92-02.

Marshall, J. 1991. A review of methods for assessing the sensitivity of aquifers to pesticide contamination. In (D.L. Weigmann, Ed) Pesticides in the Next Decade: The Challanges Ahead, Proceedings of the Third National Research Conference on Pesticides, NOvember 8-9, 1990. Virginia Water Resources Research Center, Virginia Polytechnic Institute and State University, Blacksburg, 1991. pp 618-633.

Sacha, L., D. Fleming, and H. Wysocki. 1987. Survey of pesticides used in selected areas having vulnerable ground waters in Washington State. U.S. Environmental Protection Agency, Epa/910/9-87/169. Seattle, WA.

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.

Teso, R.R., T. Younglove, M.R. Peterson, D.L. Sheeks III, and R.E. Gallavan. 1988. Soil taxonomy and surveys: classification of areal sensitivity to pesticide contamination of groundwater. Journal of Soil and Water Conservation, 43:348-352.

Wilkerson, M., R.J. Oshima, T. Younglove, H. Margolis, and S.J. Marade. 1985. Pesticide movement to ground water EPA grant #009155-79. Volume III. Use of agronomic variables to predict ground water contamination in the San Joaquin Valley, California. Environmental Hazards Assessment Program, California Department of Food and Agriculture, Sacramento, Ca.