Preliminary copy: Not for publication

A-045-IDA FILE COPY 5

PAPER NO. 76-2085

FROZEN GROUND PROBABILITIES USING DISCRIMINANT ANALYSIS

by

En Sheng Yen
Department of Forestry
Ping Tung Institute of Agriculture
Taiwan
Republic of China

Myron Molnau
Ag. Engineering
University of Idaho
Moscow, Idaho

D. K. McCool
Palouse Conservation Field Station
Agricultural Research Service - USDA
Pullman, Washington

For presentation at the 1976 Annual Meeting
AMERICAN SOCIETY OF AGRICULTURAL ENGINEERS

University of Nebraska
Lincoln, Nebraska
June 27-30, 1976

SUMMARY: Floods occurring on frozen ground are destructive, both from the point of view of flood severity and erosion hazard. Discriminant analysis was used to classify past runoff events as occurring on frozen or unfrozen ground. For the Greater Palouse area, the average minimum temperature during the freeze period was the single most important variable in classification of events. The method did not appear to work well for a large basin in southern Idaho.

American Society of Agricultural Engineers

St. Joseph, Michigan 49085

INTRODUCTION

Frozen ground is one of the important factors causing winter floods in the Pacific Northwest during the winter and early spring months. On February 26, 1948, a flood in the South Fork Palouse River Basin caused extensive damage in Pullman, especially in a mobile home court and the central business district, both of which are located on the only flat land in town in the flood plain. This flood was probably triggered by rainfall on frozen ground. In 1962 and 1964, there were two other very severe floods in Idaho. The USGS and SCS reported that the cause of flooding was prolonged rainfall on snow and frozen soil. Johnson and McArthur (1973), in a study of winter flooding in Idaho and surrounding areas for 1955-1972, found that the amount and intensity of the rainfall, the amount of snowmelt, and the imperviousness of the frozen soil combined to affect the flood severity. Virtually every large January flood in southern Idaho is a frozen ground flood. Because of extensive damage from each flood, millions of dollars have been spent on flood control every year.

In a 33 year record, from the erosion season of 1939-40 through the erosion season of 1971-72, over 40 percent of the soil loss in Whitman County, Washington, was associated in some way with rain on frozen soil or rain on melting snow. These major soil loss events, of course, did not cause all of the soil loss for a given year, but they frequently set the pattern for that which was to follow.

OBJECTIVES

The overall objective of this project was to obtain some information on the probability of the ground being frozen when a runoff event occurs. In order to do this, several specific goals were set.

1. Identify the most readily available meteorological variables that are important in causing frozen ground.
2. Develop a procedure whereby past runoff events which occurred on frozen ground could be identified.

PROCEDURE

Background

Simple techniques that use readily available climatological data in predicting both the frequency and severity of frozen ground conditions would be more useful than more complicated procedures, especially if used in flood forecasting procedures.

Numerous formulas have been developed for predicting the depth of frost penetration. Two of the easiest to use are the Stefan formula and the Berggren equation (Aldrich, 1956). The Stefan formula uses a surface freezing index which is the number of degrees deviation of the mean daily soil surface temperature from 32° F. It also includes the thermal conductivity of the soil. The Berggren equation uses the soil surface temperature directly and employs a coefficient to make a correction for the volumetric heat of fusion of the frozen and unfrozen soil. Although both of these formulas use surface temperatures, such data are virtually non-existent so that air temperature is normally used to index both the freezing index and surface temperature.

Frost penetration and thaw equations generally work much better in regions where the soil is continuously frozen throughout the entire winter. In regions of intermittent freezing and thawing, the various equations are difficult to apply at best. According to Brown (1964), regions with a total annual freeze index of less than 100 can be expected to exhibit no correlation between accumulative freeze index and depth of frost because occasional cold snaps of a few days overshadow any seasonal effects.

In an area such as the greater Palouse, the weather is such that freeze-thaw cycles play an important part in the erosion and flooding problem. The soil seldom, if ever, freezes for the entire winter; instead freeze-thaw cycles occur. Hershfield (1974) presented graphs which show an average of 10 and 20 cycles for January and February respectively in the Palouse area. Southern Idaho, on the other hand, experiences about 10 cycles in each month. Each one of these cycles represents the possibility of frost being produced in the soil.

Winter flooding in Idaho was investigated by Merrell (1964) and McArthur and Merrell (1972). Merrell investigated the joint probabilities of these parameters believed associated with frozen ground floods. These parameters were daily precipitation, mean daily temperatures, and snow depth. Since soil moisture before freezing has a significant effect on the infiltration capacity of the frozen soil, he also included the monthly temperature and precipitation of the previous September through December as an index of antecedent soil moisture. Based on his study, he concluded that the following three parameters could be used as indicies of possible flood producing conditions: 1) daily snow melt plus rain indexed water available for runoff, 2) October through December total precipitation indexed the antecedent soil moisture and 3) October through December temperatures indexed the amount and extent of soil frost. McArthur and Merrell (1972), in a study of two watersheds, found that high autumn precipitation and low autumn temperatures are usually associated with winter flooding in southern Idaho.

Variables Used in the Present Study

Based on the literature, several variables were selected for further study. These variables are shown in Table 1. Each variable may be classed into one of several groups. Since Bloomsburg and Wang (1969) showed that the moisture content of the soil upon freezing is important to the permeability of the frozen soil, indicies of antecedent soil moisture are needed. These are variables 10 and 11. The severity of the freeze is indicated by variables 1, 2, 7 and 12 while the water available for runoff is indexed by 4, 5 and 6. The actual runoff is variable 8, which is used to indicate either the severity of flooding or alternatively, an indication as to whether the runoff is greater than expected for the rain plus melt. Item 9, the thaw index is used to indicate soil warming; in other words, how much heat is necessary to later cool the soil to the point where it would again freeze. In most cases, more reasonable results were obtained if the snow parameters were simulated rather than using published values.

TABLE 1. VARIABLES USED IN FROZEN GROUND ANALYSIS

	Variable	Unit
1.	Number of days in freeze period	days
2.	Accumulative freeze index	degree(F)days
3.	Precipitation 2 days before the start of freeze period	inches
4.	Precipitation during freeze period	inches
5.	Precipitation 4 days after freeze period	inches
6.	Snowfall during freeze period	inches
7.	Depth of snow on ground at start of freeze period	inches
8.	Streamflow for 4 days after freeze period	cfs-days
9.	Accumulative thaw index	degree(F)days
10.	Precipitation from 1 Oct to start of freeze period	inches
11.	Number of days from 1 Oct to start of freeze period	days
12.	Average minimum temperature during freeze period	(F)

Discriminant Analysis

Discriminant analysis is a multivariate statistical technique that can be used to 1) determine the important variables that distinguish observations or individuals in one group from observations or individuals in other groups and 2) on the basis of these variables, assign future observations to a particular group. A group of variables or characteristics is used on which the various groups are expected to differ. In the present case, the two groups are frozen or unfrozen ground runoff events and the characteristics chosen are those in Table 1. Another grouping could also be simply whether or not the soil is frozen. The objective of this type of analysis is to obtain a function or functions which is a linear combination of the selected variables. These functions then give a single number that can be used to classify future observations into one of the groups and can be used to assign a probability to the correctness of classification. A full discussion of the procedure is found in Yen (1975) and DeCoursey (1973).

Several criteria must be met by a discriminant analysis program. The first is that the program must be able to handle unequal prior probabilities. In most cases, there is not an equal probability that any new observation will be assigned to any one of the groups. Usually, there is a greater probability that the new observation will go into one group than another. In the present case, there is a greater probability that any new runoff event will occur on unfrozen ground than on frozen ground.

The second criterion concerns unequal covariance matrices. Each data group has a covariance matrix. If the matrices are not equal, then different techniques must be used for classification than if they are equal. (Anderson and Bahadur, 1962)

The programs used in the present analysis meet both of the criteria.

EXPERIMENTAL AREAS

Four watersheds were chosen for study on the basis of the flooding problems in the basin, the availability of data (particularly good climatic data) and the number of freeze-thaw cycles the basin may experience in a given winter. The basins chosen were Missouri Flat Creek near Pullman, Washington, McKay Creek near Pilot Rock, Oregon, Blackfoot River near Blackfoot, Idaho and Rock Creek near Rock Creek, Idaho (Figure 1).

Missouri Flat Creek is a 27.1 sq. mile watershed averaging 2800 feet in elevation. The nearest climatic station is Pullman, Washington, one mile outside the basin. The mean annual maximum temperature is about 57^{o} F and the mean annual minimum is about 36^{o} F. The area experiences about 10 freeze-thaw cycles in January and 20 cycles in February (Hershfield, 1974). During November through March, the monthly precipitation is over 2 inches per month with January having 3.13 inches and a mean temperature of 28^{o} F.

McKay Creek near Pilot Rock, Oregon, has an area of 180 sq. miles. The nearest climatic station is Pendleton, Oregon, about 30 miles outside the basin. The mean annual maximum and minimum temperatures are 63^{o} F and 42^{o} F respectively, uncorrected for elevation. The freeze-thaw pattern is approximately the same as for Missouri Flat Creek. During October through March, each month has over one inch of precipitation with December and January having equal amounts of 1.56 inches and January having a mean temperature of 32^{o} F.

The Blackfoot River is the largest of the basins chosen with a drainage area of 1295 sq. miles and a mean elevation of about 5800 feet. The weather station used with this basin was Aberdeen, Idaho. The mean annual maximum and minimum temperatures are 60° F and 31° F, respectively with a December precipitation of 0.80 inches and a mean temperature of 25° F. The precipitation is relatively well distributed throughout the year with only May having more than one inch (1.08 inches) and only July and August having less than 0.5 inches (0.42 and 0.47 inches). All of southern Idaho has about 10 freeze-thaw cycles in each winter month.

Rock Creek will not be reported on in this report since the data analysis is not yet complete.

RESULTS AND DISCUSSION

In the first part of this study, the freeze index, precipitation and streamflow (variables 2, 5 and 8, Table 1) were used in a preliminary analysis. From investigation of the Missouri Flat Creek data for 1953 through 1973; it was evident that individual freeze events with freeze index values of less than 100 were generally not associated with larger than normal runoff. Values of 150 freezing index units and over seldom left any doubt that the infiltration capacity of the soil was severely impaired because the runoff was much greater than expected for the amount of rain and snowmelt for unfrozen soil.

It is, of course, possible for high rates of runoff to occur independently of frozen soil. This occurs principally from snowmelt in higher elevations. The gaging station is at a lower elevation so that unless the air temperature and precipitation records are checked closely, it may be erroneously assumed that high observed runoff occurred because of rainfall on frozen ground.

Based on the results of the preliminary Missouri Flat Creek study, it was felt that climatic and hydrologic variables could be used to provide a method for sorting or classifying past runoff events as to whether or not they occurred on frozen ground. Discriminant analysis is a method that can be used for classification studies of this type.

Missouri Flat Creek

The full set of data listed in Table 1 was collected for Missouri Flat Creek. A total of 143 runoff events were chosen for analysis. From a detailed analysis of these events, 54 could be classified as frozen or unfrozen ground runoff events, leaving 89 unknown cases.

The first eight variables of Table 1 were used in an attempt to use a large amount of information for classification. The output of a discriminant analysis program consists of a linear function for each group. Also, a canonical function for each group can also be computed (Dixon, 1975). These functions essentially give the correlation between the two discriminant functions. They can also be plotted and will give an excellent graphical portrayal of the degree of separation of the two groups.

Figure 2 shows this grouping of observations from a preliminary test about two centroids. Since the probability of classification into one group or the other is inversely proportional to the distance of the observation from the centroid, this plot gives a good illustration of the scatter of the data. The distance between group centroids is a good indication of the strength of the classification. The misclassified events are quite prominent in such a plot.

When the 54 known cases of frozen and unfrozen ground events were run through the discriminant analysis program, the program derived the necessary functions and then checked all 54 observations for correct classification. The program reclassified 2 of 12 events which had been entered as frozen ground events to unfrozen ground events and 1 of 42 was reclassified as frozen from unfrozen ground events (Figure 3).

The 89 unknown observations were then subjected to the analysis using the same functions as derived from the known data. Ten of the unknown cases were classified as frozen ground runoff events while the other 79 were classified as unfrozen ground runoff events. The probability of the individual events being correctly classified ranged from a low of 53 percent to a high of 100 percent with the majority being in the 90's.

The next investigation undertaken was to determine which of the variables used above would be the best ones to use in the classification. A stepwise discriminant analysis procedure was used (Dixon, 1975). The step-wise selection process of the 8 variables resulted in the use of only two, the total freeze index and the flow 4 days after the event. Using these two variables, 93 percent of the known events were correctly classified as opposed to 94 percent when using all 8 variables (Figure 4). However, these two variables classified only 7 of the 89 unknown cases as frozen ground events as opposed to 10 when using 8 variables nor were they all the same events. It appears that the classification equations which contain only two variables contain significantly less information than the set of 8 variables and thus cannot discriminate those cases for which the probabilities are near 50 percent.

TABLE 2. STEPWISE SELECTION OF VARIABLES

Variable (Table 1)	Missouri Flat Creek	McKay Creek	Blackfoot River
1			1
2	6		
3			
4	5	2	
5			4
6	3		
7			
8	2	3	
9	4		
10		4	
11			2
12	1	1	3

When all 12 of the variables were used in the step-wise selection procedure, only 6 variables had a F value of 1.00 or higher which was the F value used as the level at which further variables were not significant. Table 2 lists the order of selection of variables for all 3 study streams. Thus it is seen that for Missouri Flat Creek, the variable with the highest F value was variable 12, the average minimum temperature during the freeze period and the variable with a F value nearest to but greater than 1.00 is variable 2, the freeze index. Figure 5 shows the canonical plot of Missouri Flat Creek when using these 6 variables.

A comparison of the variables selected for all three basins reveals that the average minimum temperature was selected as the most important variable in all cases except for the Blackfoot River where it was the third selected variable. It would appear that the average minimum temperature contains enough information to indicate whether or not the ground is frozen. For all of these watersheds, the fact that the average minimum temperature is low is enough to insure that a frost layer exists and would inhibit infiltration. The depth of penetration would not be important if it is only desired to determine the fact of reduced infiltration. The depth of penetration is included in the freeze index or the average minimum temperature plus the number of days the temperature is below freezing.

McKay Creek

For McKay Creek, 70 runoff events were isolated, 12 of which were thought to have occurred on frozen ground. Figure 6 shows a plot of the canonical functions for selected events only. It can be seen that these events are well separated and it can be concluded that these four variables (12, 4, 8 and 10) can be used to classify past frozen and unfrozen ground runoff events for McKay Creek.

It can be seen from Table 2 that for McKay Creek, the precipitation from the start of the water year to the freeze date is important. This probably is because of the dry nature of the basin and if soil moisture is not high, even frozen soil will not exhibit reduced infiltration. The soil could be frozen but if the infiltration capacity were not reduced the statistical procedure would not consider it to be frozen ground.

Blackfoot River

For the Blackfoot River basin in southern Idaho, it was quite difficult to determine, just from the data on hand, whether the ground was frozen or not. Therefore, soil temperature records were used to indicate the presence of frost. If the maximum temperature were 32^{o} F or less at a depth of four inches, the soil was classified as frozen. Using this definition of frozen soil, 54 frozen ground events out of a total of 83 observations were identified. When using the step-wise discriminant analysis procedure, the four variables shown in Table 2 all had a F value greater than 1.00. This is a big watershed; it is also dry and cold. From Table 2, the number of days in the freeze period probably indicates that it takes several days for the entire basin to be frozen over to the point where the basinwide infiltration capacity is impaired. Also, the large area would most likely not contribute runoff all at the same time. The number of days from October 1 to the freeze date would be an index of the soil moisture as well as the temperature itself which is also represented by the average minimum temperature. The streamflow is not explicitly stated but is indexed by the precipitation for four days after the freeze period is over.

By using these four variables, only 74 percent of the observations were correctly classified. Figure 7 shows the plot resulting from the classification. It is evident that these are quite scattered and are not clustered as well as the values in Figures 5 and 6. There is much overlapping and the centroids of the groups are close together, making it very difficult to visually separate them.

Probabilities

For each of the classified events, the probability that the event is a frozen or unfrozen ground runoff event can be computed. This is done by using a conditional probability model (Barr and Goodnight, 1972). Thus for each observation, the probability can be calculated that the observation belongs to one group or the other. Then if the probability is over some set amount, say 50 or 60 percent, the observation could be put into that group. Overall and Klett (1972) go into this problem in some detail. In the present study, few problems with classification were found other than lack of separation. For the majority of the cases, standard statistical packages available at most computing centers should prove adequate.

CONCLUSIONS

The main purpose of this study was to investigate the use of discriminant analysis in determining whether or not a runoff event occurred on frozen ground. It would appear that the success of the method depends upon the size, climatic regime and location of the watershed.

The procedure appeared to work well for Missouri Flat Creek, a relatively small watershed with pronounced freeze-thaw cycles. The method also worked well, although less so, for McKay Creek, a larger drier area. For the much larger Blackfoot River Basin in southern Idaho, the method did not appear to give usable results probably because of the large size of the basin and the lack of pronounced freeze-thaw cycles.

The average minimum temperature during the below freezing period was important to the classification procedure in all three basins and ranked first for two of the basins. The number of days below freezing was the most important variable for one basin.

One of the outputs from a discriminant analysis program is the probability of an observation correctly classified.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Office of Water Research and Technology for financial help through the allotment program of the Idaho Institute for Water Resources Research. The Agricultural Research Service at Pullman, Washington also provided guidance and supplied the data tape used in the analysis. The overall guidance of the project was in the Agricultural Engineering Department, University of Idaho. During this study, Mr. Yen was on a sabbatical leave supported by the government of the Republic of China. Mr. Ralph Pedersen, Agricultural Engineering Department did the calculations for McKay Creek and Blackfoot River.

LITERATURE CITED

Aldrich, Karl P., Jr. 1956. Frost Penetration Below Highway and Airfield Pavements. Highway Research Board Bulletin 135. pp. 124-144.

Anderson, T. W. and R. R. Bahadur. 1962. Classification into Two Multivariate Normal Distributions with Different Covariance Matrices. Annuals of Mathematical Statistics 33(2): 420-426.

Barr, Anthony James and James Howard Goodnight. 1972. A Users' Guide to the Statistical Analysis System. North Carolina State University.

Bloomsburg, G. L. and S. J. Wang. 1969. Effect of Moisture Content on Permeability of Frozen Soil. Paper presented at the 16th Annual Meeting, Pacific Northwest Region, American Geophysical Union, Portland, Oregon.

Brown, W. G. 1964. Difficulties Associated With Predicting Depth of Freeze or Thaw. Canadian Geotechnical Journal 1(4): 215-226.

DeCoursey, D. G. 1973. Application of Discriminant Analysis in Design Review. Water Resources Research 9 (1):93-102.

Dixon, W. J. 1975. BMDP, Biomedical Computer Program. University of California Press, Berkeley. 792 pp.

Hershfield, David M. 1974. The Frequency of Freeze-Thaw Cycles. Journal of Applied Meteorology 13:348-354.

McArthur, Robin and J. Evan Merrell. 1972. Flood Runoff From Rainfall and Snowmelt Over a Frozen Soil Mantle. ARS-SCS Snow Hydrology Workshop. Boise, Idaho.

Merrell, J. Evan. 1964. Computing Joint Probabilities Associated with Combination Rainfall, Snowmelt, and Frozen Ground Conditions. Unpublished Mimeo, Soil Conservation Service, Boise, Idaho.

Nie, Norman H., C. Hadlai Hull, Jean G. Jenkins, Karin Steinbrenner and Dale H. Bent. 1975. SPSS, Statistical Package for the Social Sciences (2nd Edition). McGraw-Hill Book Company. 675 pp.

Overall, John E. and C. James Klett. 1972. Applied Multivariate Analysis. McGraw-Hill Book Company, New York.

Yen, En Sheng. 1975. The Determination of Frozen Ground Probabilities from Climatic and Hydrologic Data. Master of Science Thesis. Agricultural Engineering, University of Idaho, Moscow.

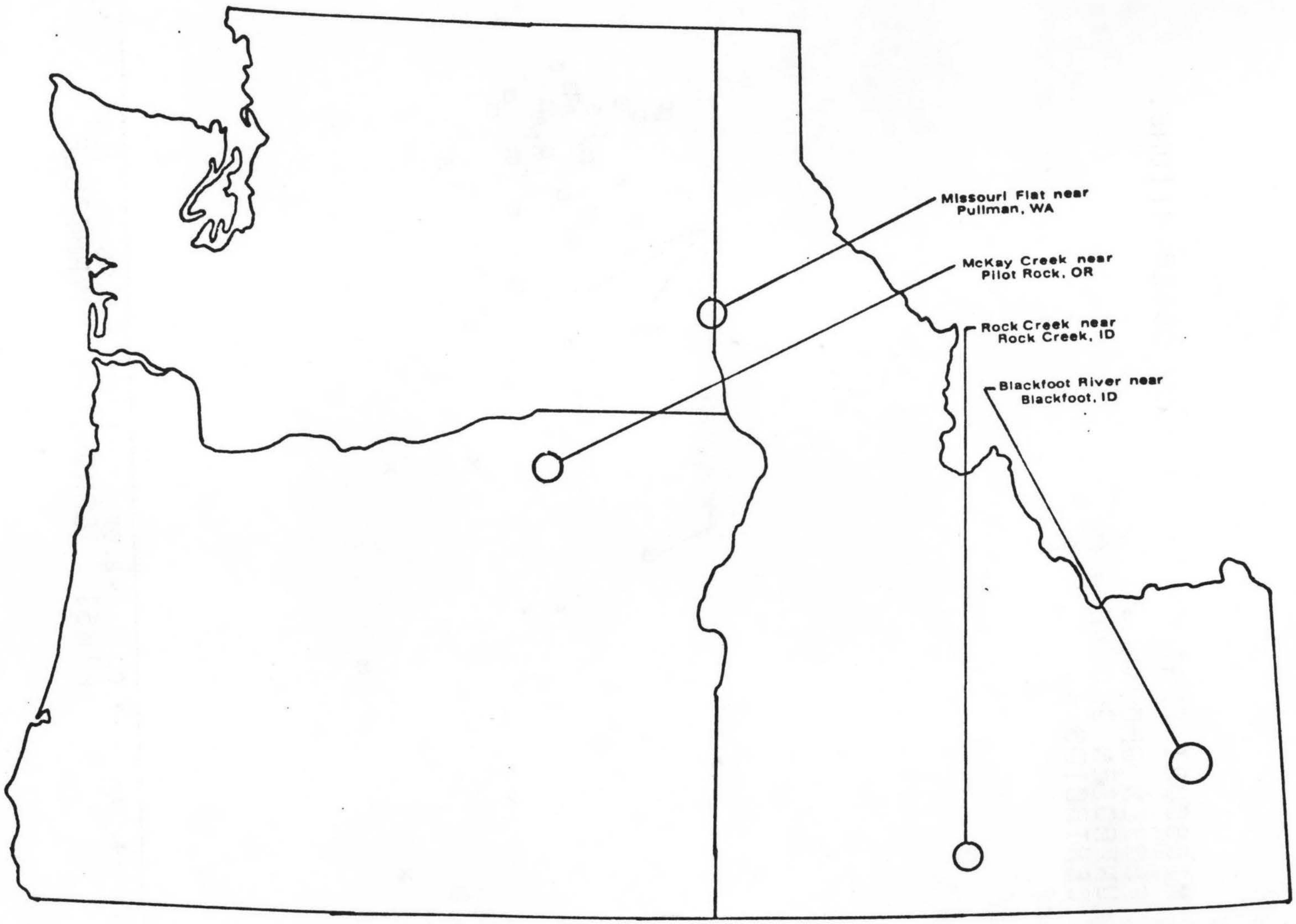

Figure 1. Location map showing study sites.

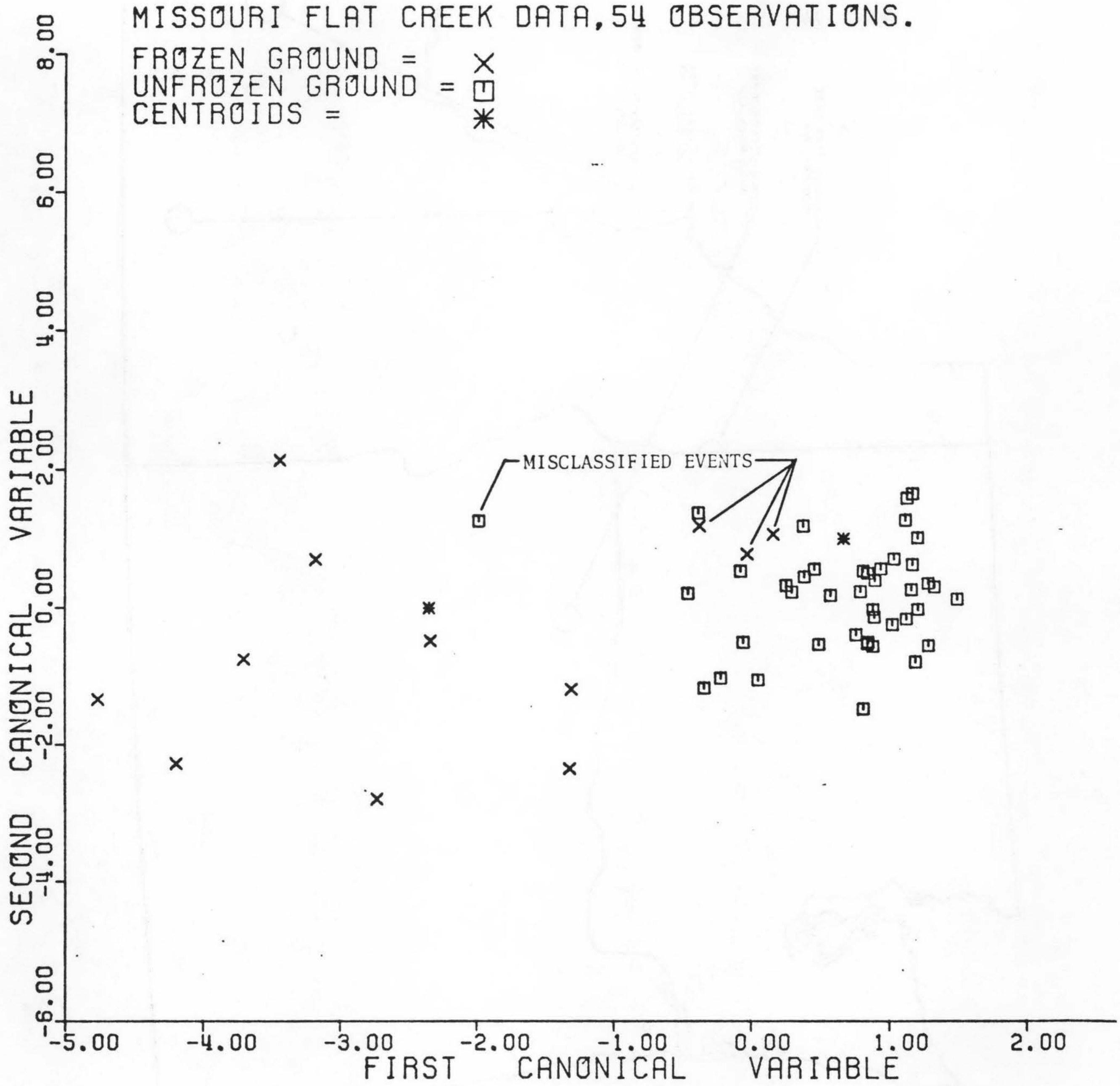

Figure 2. An example of the clustering of frozen and unfrozen ground runoff events.

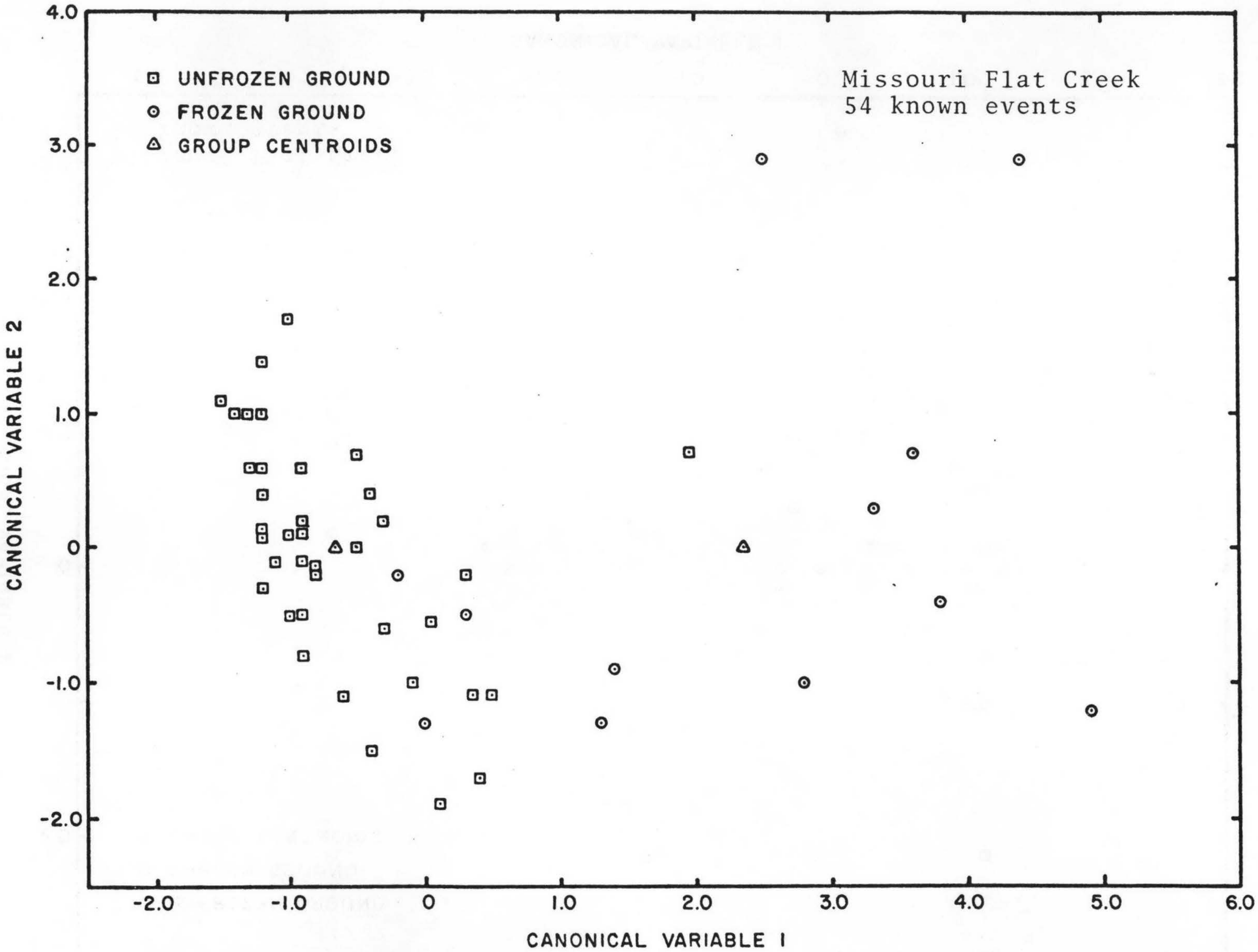

Figure 3. Classification using variables 1 through 8.

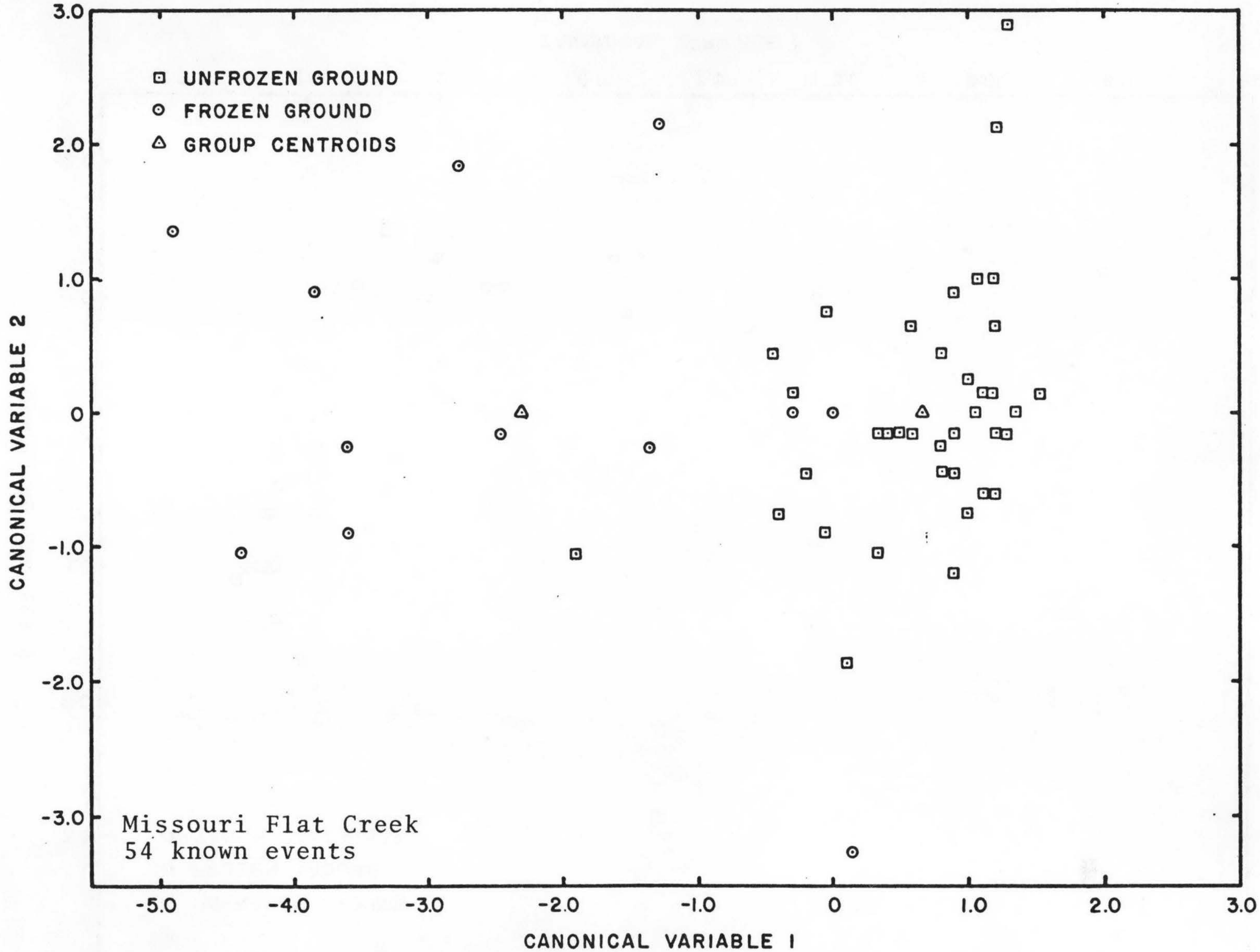

Figure 4. Classification using variables 2 and 8.

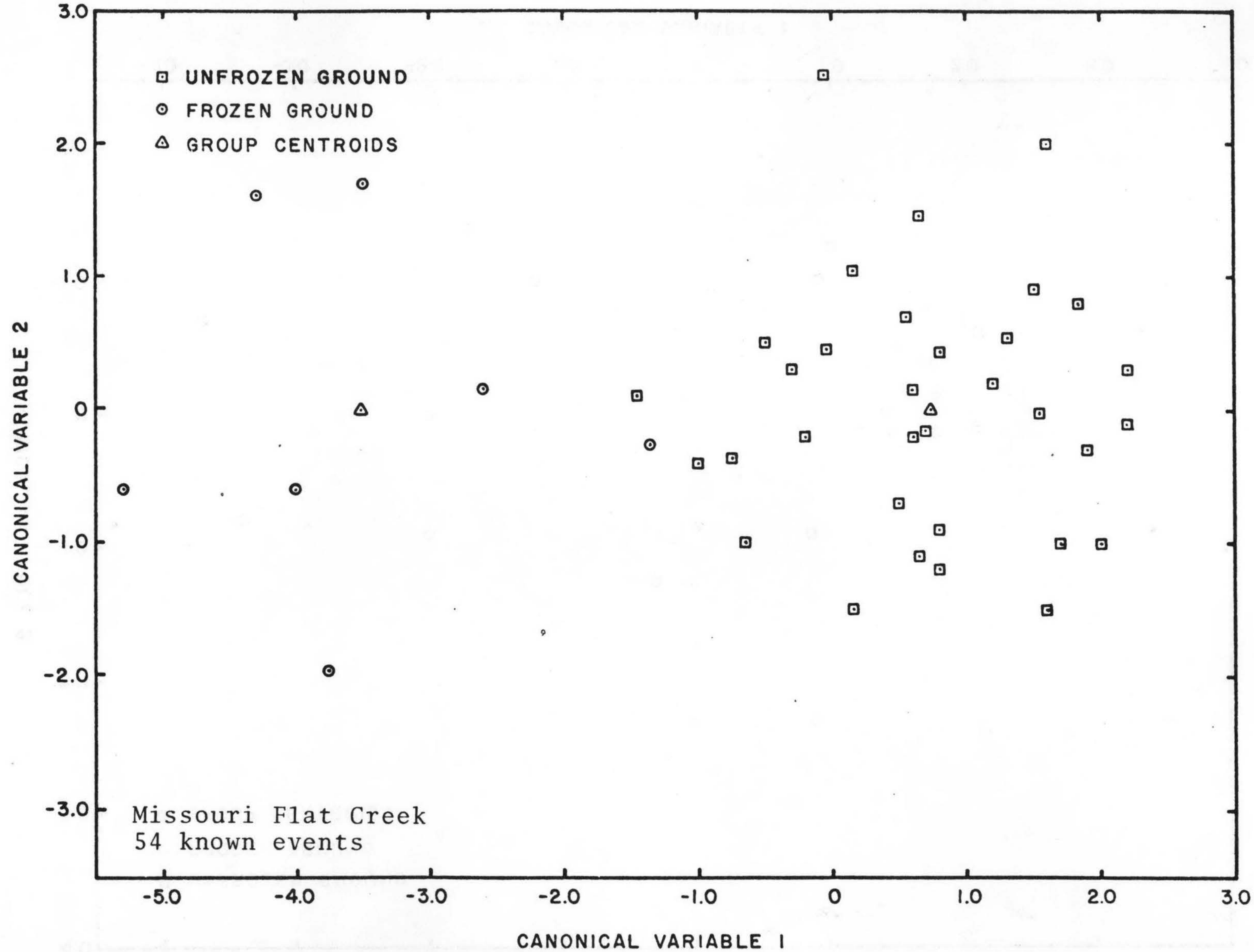

Figure 5. Classification using variables 2, 4, 6, 8, 9 and 12.

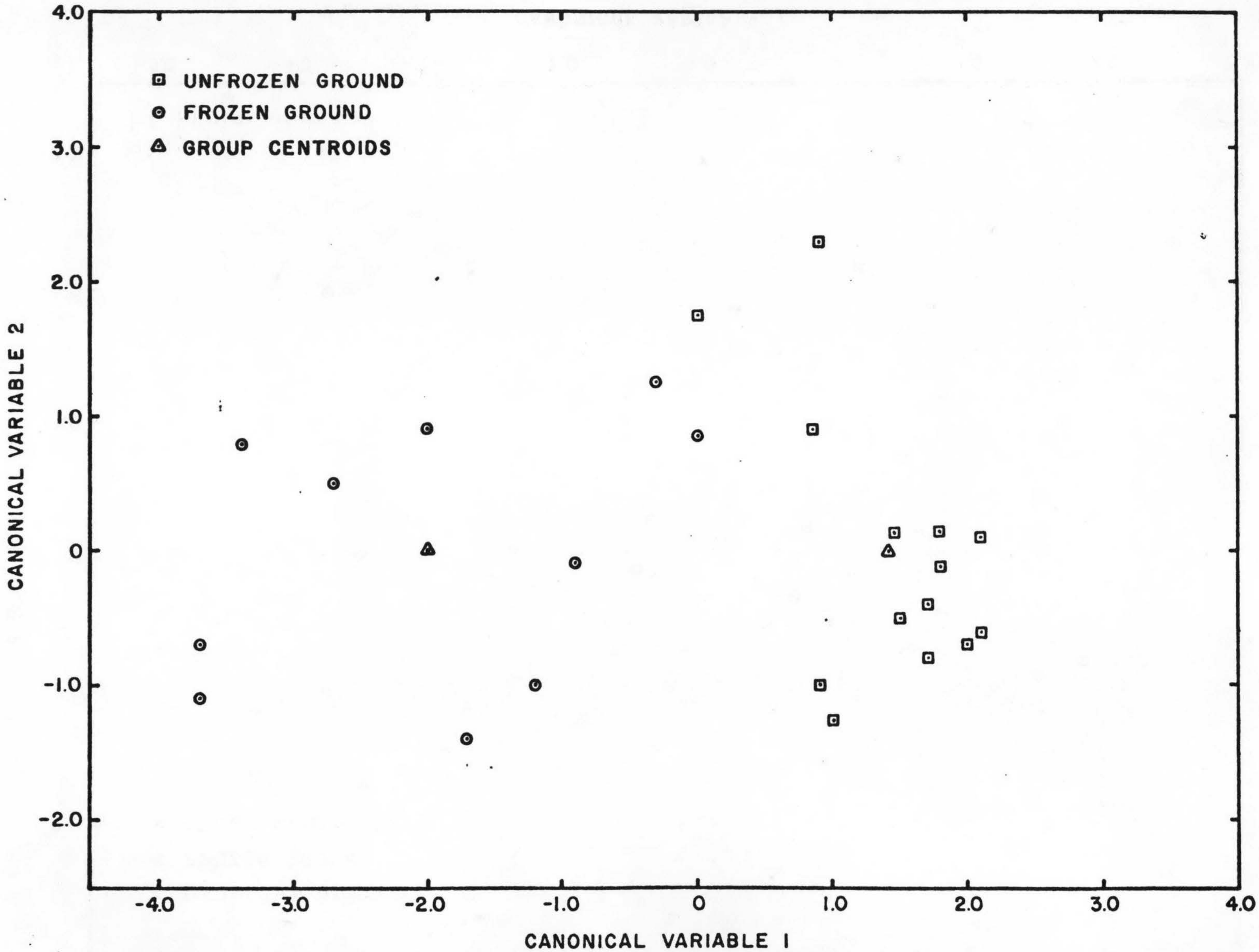

Figure 6. Classification of selected events for McKay Creek.

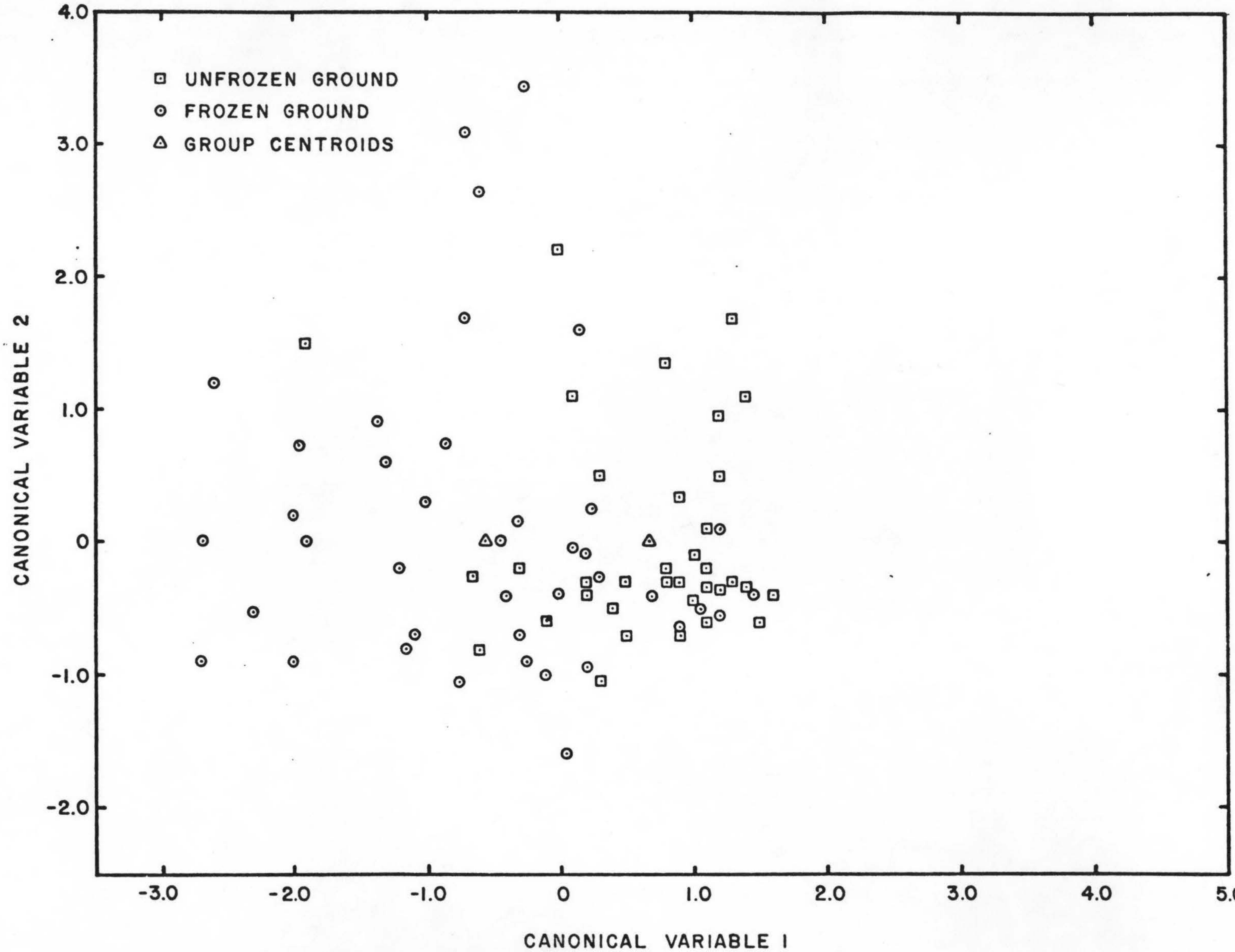

Figure 7. Classification of runoff events for the Blackfoot River.

Figure 7. Classification of [illegible] for the [illegible]