APPLICATIONS OF CLASSIFICATION TREE METHODOLOGY TO AVALANCHE DATA MANAGEMENT AND FORECASTING

Robert E. Davis,[1] Kelly Elder [2] and Eddy Bouzaglou[3]

ABSTRACT

A common problem for avalanche-prone areas subject to forecasting and control responsibility is the identification of a set of critical meteorological parameters. While storm snowfall and snowfall intensity are generally accepted as among the most important, the complex interplay between terrain, wind, temperature, solar radiation and other meteorological variables makes identifying the next most important parameters difficult. Classification tree methodology is introduced as a potential tool for identifying critical meteorological parameters associated with avalanche and control activities. The application of this methodology is described in the context of exploring a subset of the Mammoth Mountain avalanche and meteorology data base. The meteorological parameters most important to avalanche occurrence in two-years of data were identified from a set of thirteen variables from one observation site. It is shown how this information could be used to provide inputs to forecasting programs and guidance in establishing observation priorities.

INTRODUCTION

It has been a long standing problem in avalanche forecasting to select a set of parameters sensitive to avalanche hazard and activity, which can be measured frequently in an operational setting. Another aspect of this problem is to establish priorities in an operational observation program ensuring that critical variables are consistently measured. Researchers have proposed and described different avalanche climates for various mountain regions for several decades. These mostly descriptive works (c.f. LaChapelle, 1965; Armstrong and Armstrong, 1987) suggest that within specific regions a limited set of meteorological variables is usually responsible for avalanche activity. However, it may not be possible to identify a consistent set of parameters for all types of avalanches, storms and triggering mechanisms in the scope of avalanche forecasting for large regions.

Primary variables and trends in these variables associated with the occurrence of avalanches have been identified in local-area studies (Judson and Erickson, 1973). Examples of this work include the extensive investigations carried out by researchers at the Swiss Federal Institute of Snow and Avalanche Research (SFISAR) (e.g. Fohn et al., 1977; Obled and Good, 1980; Buser, 1983, 1989; Buser and Good, 1984; Buser et al., 1987). Analyses of avalanches in small areas may be aided by unambiguous and reliable avalanche occurrence reporting, a major problem for evaluation on a regional basis. However, results from local studies are not generally applicable to other areas. For example, snow covered slopes at ski areas, over highways and in other situations where avalanches are artificially released may have different responses to meteorological variables than similar slopes where only naturally triggered avalanches run. Natural avalanche slopes are also affected by different loading and triggering

[1] Geophysical Sciences Branch, U.S. Army Cold Regions Research and Engineering Laboratory, 72 Lyme Road, Hanover NH, 03755

[2] Center for Remote Sensing and Environmental Optics, University of California, Santa Barbara, CA 93106

[3] Ski Patrol, Mammoth/June Ski Resort, P.O. Box 24, Mammoth Lakes, CA 93546

mechanisms, enough so that Judson (1983) suggested that index slopes yield little diagnostic information in a regional context.

Nevertheless, in areas where the slopes are controlled and avalanche observations are reasonably complete, recently developed methods can provide useful avalanche forecast and control guidance. An example of this is the Swiss nearest neighbors method, implemented in the software package NXD developed at SFISAR. This method is based on the assumption that within a given area, the same meteorological conditions with the same snow conditions will produce the same avalanche activity. A set of critical meteorological and snow parameters are input to the NXD program, and the ten most similar days are selected from the historical record. The descriptions from the record have been shown to be useful to patrollers (Buser, 1989), who can compare their opinions for the day, and check for unusual activity with similar circumstances. A crux to using this method is that a set of critical parameters must be selected and, if justified, assigned weights in the program. In using NXD operationally, the critical parameters have been chosen "from 20 years of experience" (O. Buser, personal communication), but this experience has been the subject of several statistical analyses and evaluations (Fohn et al., 1977; Obled and Good, 1980). Therefore we propose here a non-parametric technique that ranks available measurements and assigns relative scores to their use in classifying avalanche activity.

The long-term goals of this study are to investigate the use of classification-tree methodology to identify a set of critical weather variables measured at Mammoth Mountain in terms of classifying avalanche activity, and evaluate both the decision tree method and the Swiss NXD nearest neighbor software for forecast and control guidance.

APPROACH

This paper reports on the preliminary application of classification-tree methodology (Classification and Regression Trees: CART™; Breiman et al., 1984) to identify critical parameters in the weather and avalanche data sets from Mammoth Mountain, California. Mammoth Mountain lies in the eastern Sierra Nevada, and occupies an elevation range from about 2,590 to 3,371 m. The data from only two winters, 1989-1990 and 1990-1991, were analyzed; a more complete analysis will be possible when the entire data record is entered into the computer database. Meteorological variables are recorded at several sites on Mammoth Mountain, at the lodges, at a U.S. Forest Service snow plot and at a University of California study site. In this study we used data from the Main Lodge at an elevation of 2,743 m on the northern base of the mountain. Variables selected from the weather record for this analysis are listed in Table 1.

Table 1. Input data used in CART analysis from daily data record.

1)	Total snow depth
2)	Storm snow depth
3)	New snow depth
4)	Snow water equivalent - storm
5)	Fractional density
6)	Average wind speed
7)	Maximum wind gust speed
8)	Maximum 24-h air temperature
9)	Minimum 24-h air temperature
10)	Current air temperature
11)	Snow surface temperature
12)	Snow temperature at 10 cm depth
13)	Snow temperature at 20 cm depth

Control activities and avalanche observations are recorded at Mammoth Mountain in a format consistent with the standard U.S. Forest Service avalanche control and occurrence chart. This protocol consists of codes for the date, time, path, patroller identification, control type, control number, control surface, avalanche class type (hard slab, soft slab, etc.), avalanche trigger mechanism, avalanche size, and so forth (Perla and Martinelli, 1978). The response variable in this analysis was a categorical parameter representing a combination of control and avalanche activity (Table 2).

Table 2. Output classes used in CART analysis of weather and avalanche data.

Class 1:	No avalanche or control activity
Class 2:	Control day, but no avalanches
Class 3:	Control day with sloughs and/or avalanches

Because there remains much ambiguity about how to assign a degree to the intensity of avalanche activity, we simply chose to classify all days with notable point releases or avalanches as avalanche days. This method provides a conservative category representing notable avalanche activity, and is consistent with the way the nearest neighbors method handles avalanche activity. All days in the record with natural avalanches were also control days and considered as Class 3.

The weather and avalanche data were analyzed using a hierarchical or decision tree classifier, CART (Breiman et al, 1984). In a study comparing the performance of neural networks and classification trees, Atlas et al. (1990) showed that in most cases the two methodologies exhibited comparable accuracy for prediction. The classification tree arrives at decisions concerning class splits through a hierarchy of stages. The main advantage of this method is that while the techniques used to construct the classification are complex, the decision tree is easy to understand and provides a straightforward procedure to predict the results. Figure 1 shows a conceptual decision tree that illustrates some of the features of this methodology.

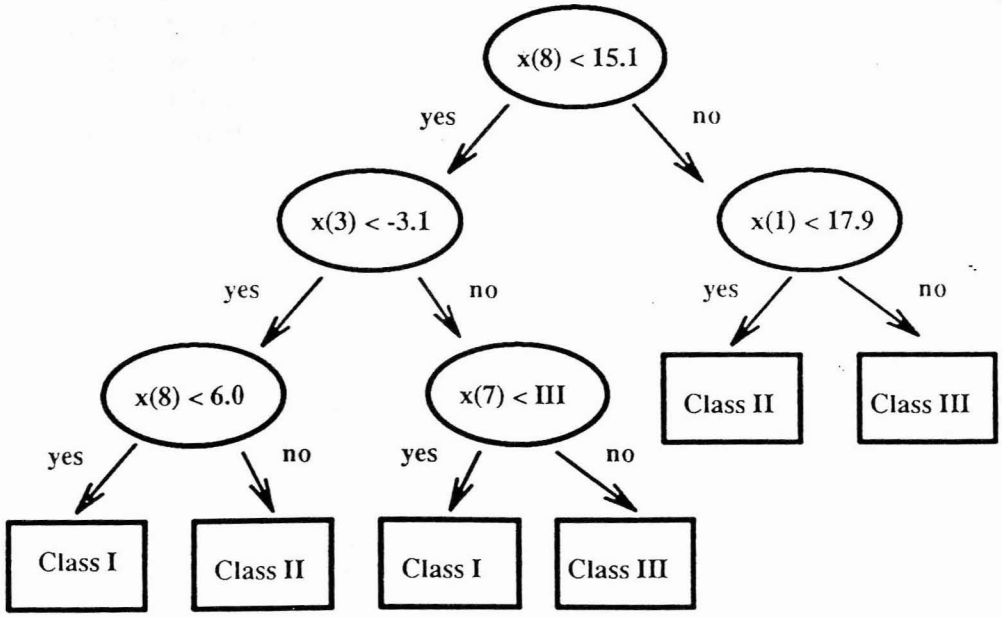

Figure 1. Conceptual classification tree.

A decision tree is composed of an ordered set of nodes, each of which is represented by a threshold decision function, or in the case of end nodes, an end partition, or class. The importance of a particular variable depends on where in the classification hierarchy the splits are made, as well as what the surrogate role of the variable is. In Figure 1 for example, the variable x(8) appears at two stages in the decision hierarchy. In a prediction mode, the tree structure is robust in accounting for missing data, because at each node the methodology finds the optimum surrogate (replacement) split variable from other parameters. This feature makes the methodology more attractive because the scores assigned to the importance of different parameters are calculated from their rating as surrogates as well as primary variables in the decision tree. Another variable, say x(13), may not appear in the primary decision nodes, but may score high in importance because it substitutes for the primary decision variables with low extra misclassification cost.

Unlike traditional statistical techniques, decision tree classifiers do not impose restrictions on the underlying distributions of the data. Moreover, categorical data can be used. The methodology offers the option, which we did not use here, of variable assigned misclassification costs, which allows the user to specify a particular class as important, so that the algorithm will minimize the error to that class.

RESULTS

Preliminary analysis of avalanche and weather data from two winters, 1989-1990 and 1990-1991, ranked the weather variables in the order of their scores assigned by the decision tree classifier. The variable number (from Table 1), the name and the relative score are shown in Table 3.

Table 3. The ranking and relative scores of the weather and snow variables.

Variable Number	Variable Name	Relative Importance
3.	Storm snow depth	100.
6.	Fractional density, new snow	85.
11.	Current air temperature	85.
14.	Snow temperature at 20 cm depth	78.
4.	New snow depth	77.
13.	Snow temperature at 10 cm depth	73.
7.	Average wind speed	70.
10.	Minimum 24-h air temperature	57.
5.	Snow water equivalent - storm	47.
9.	Maximum 24-h air temperature	33.
12.	Snow surface temperature	32.
2.	Total snow depth	29.
8.	Maximum wind gust speed	26.

The two-year data set consisted of 380 cases (days) of which 7 were control days with no avalanches (Class 2) and 35 were avalanche days (Class 3). The rest were non-avalanche days (Class 1). The classification and classification probability matrices are shown in Tables 4a and 4b.

Table 4a. Classification matrix.

		True Class		
		1	2	3
Predicted	1	336	0	4
	2	1	7	2
Class	3	1	0	29

Table 4b. Classification probability matrix.

		True Class		
		1	2	3
Predicted	1	0.99	0.00	0.11
	2	0.00	1.00	0.06
Class	3	0.00	0.00	0.83

The decision tree is not shown for the sake of brevity and because only two years of data were analyzed. However, it may be worthwhile to note that out of 27 terminal nodes in the decision tree, 10 had only one or two cases, and all 10 were class 2 or 3. If, when the entire data set is analyzed (a few thousand days), there are likewise terminal nodes with one or two cases, then this may represent the detection of rare or outlying situations provided that there is a large population of cases in all classes. However, if these do represent rare events, they cannot be classified with much confidence because of their low number. In an operational context simple identification of unusual events may be valuable.

DISCUSSION

The classification tree methodology ranked and scored the variables (Table 3), but more than half of the variables scored 70 or higher in this data set. In this case the storm snow depth ranked highest, and by comparing the record of this variable with the avalanche activity class over time for the two winter seasons (Figures 2a and 3a) it is easy to see the association. Figure 2a shows a time series of the avalanche/control activity class and the storm snow accumulation for the 1989-1990 winter. Figure 3 shows the 1990-1991 winter.

The third ranked variable, the current air temperature, is also straightforward to explain. Winter temperatures are generally moderate in the Sierra, compared with continental ranges, and the temperature typically drops after the front has passed. Therefore, the a low temperature threshold decision in combination with the storm snow accumulation is rated high to splits classes of avalanche and control activity. Figure 2b shows a time series of the avalanche/control activity class and the current air temperature, usually observed around 9 AM, for the 1989-1990 winter and Figure 3b shows the 1990-1991 winter.

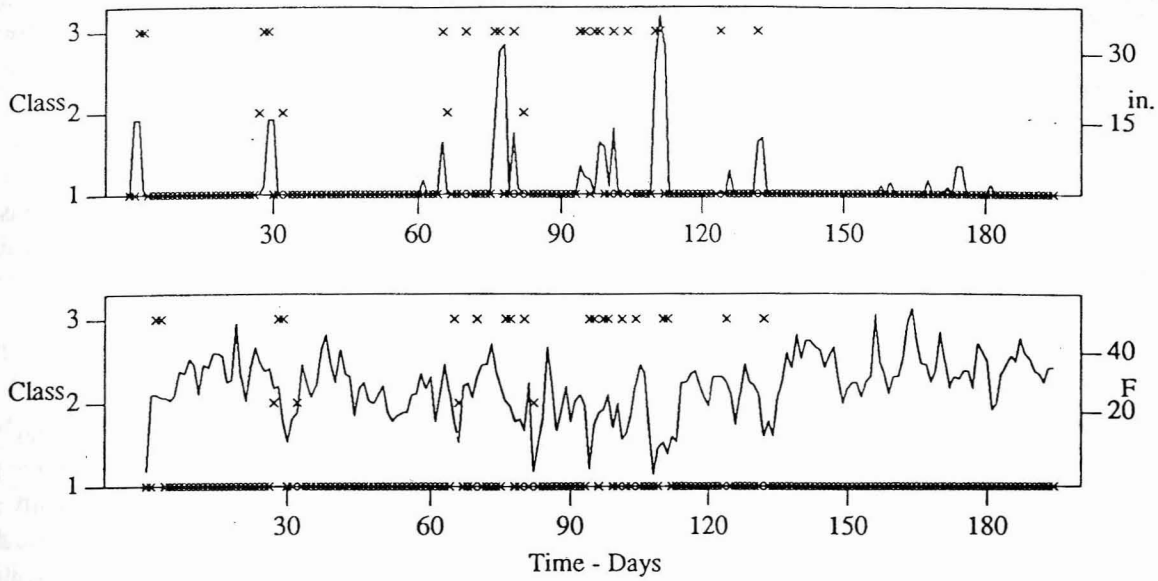

Figure 2. Avalanche activity class × plotted with total storm accumulation (line) [2a], and avalanche activity class × plotted with current air temperature (line) [2b] for the 1989-1990 snow season at Mammoth Mountain, California. The season began on October 10, 1989 (day 1) and ended May 13, 1990 (day 194).

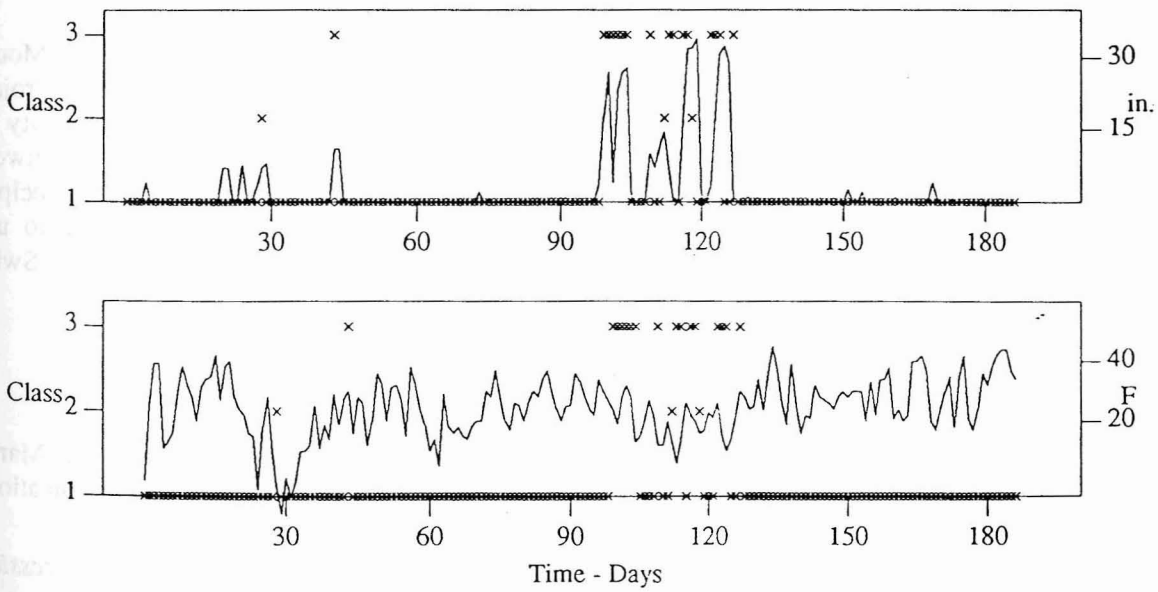

Figure 3. Avalanche activity class × plotted with total storm accumulation (line) [3a], and avalanche activity class × plotted with current air temperature (line) [3b] for the 1990-1991 snow season at Mammoth Mountain, California. The season began on November 21, 1990 (day 1) and ended May 27, 1991 (day 186).

It is expected that the ranking and scoring will be more robust once the data from several years and sites are analyzed. But even with this limited data the method shows reasonably low misclassification costs (Table 4). This result suggests that the method will be useful in providing the NXD software with a list and weights of critical variables. In addition, this method may provide important guidance to the question; what measurements are crucial to obtain?

Analysis of additional data from other sites may identify variables with different sensitivity to the classes of avalanche activity than the parameters used in this analysis. Most of the starting zones on Mammoth Mountain are above the timberline, well above the Main Lodge study plot. Other sensors and measurements are made higher up on the mountain, and those data that are accessible in real time at the ski patrol office may prove to be more useful than some of the parameters analyzed here.

CONCLUSIONS

Classification-tree methodology looks promising for identification of important weather variables for predicting a set of avalanche activity classes. The weather variables are ranked and scored by the classification, providing a list and weights that could be used in the Swiss nearest neighbor software NXD. The decision tree constructed during classification of the learning sample may provide valuable guidance to forecasting or control efforts. Preliminary results show a low overall misclassification error rate. The results are easy to interpret, and it is easy to incorporate personal bias operationally because the binary decisions are based on threshold relations, whose values can be adjusted. The next stages of this project will analyze a much larger data set, including more weather stations on Mammoth Mountain and more variables, and will adjust downward the target misclassification error rate for the avalanche category.

ACKNOWLEDGEMENTS

This work was initiated after discussions with Gary Reitman, Ski Patrol Director of Mammoth Mountain, in September, 1990. This work was funded under Army Corps of Engineers Project 4A762784AT42 with other generous support from Mammoth/June Ski Resort and the University of California. Mammoth/June Ski Resort staff, including Mary Walker, Stacey Tracy and Ron Cornwell, provided much help with understanding and transferring the data sets. Walter Rosenthal also helped with useful comments about the avalanche control records and shoot procedures. The idea to use CART to select critical variables came after detailed discussions with Othma Buser of the Swiss Federal Institute of Snow and Avalanche Research.

REFERENCES

Atlas, L., R. Cole, Y. Muthusamy, A. Lippman, J. Connor, D. Park, M. El-Sharkawi, and R. J. Marks II, 1990, "A performance comparison of trained multilayer perceptrons and trained classifications trees," *Proceedings of the IEEE,* Vol. 78(10), pp. 1614-1618.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, 1984, *Classification and Regression Trees,* Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

Buser, O., 1983, "Avalanche forecast with the method of nearest neighbours: An interactive approach," *Cold Regions Science and Technology,* Vol. 8(2), pp. 155-163. Buser, O. and W. Good, 1984, "Avalanche forecast: experience using nearest neighbours," *Proceedings, International Snow Science Workshop,* ISSW Workshop Committee, Aspen, CO, p.109-115.

Buser, O., M. Butler and W. Good, "Avalanche forecast by the nearest neighbor method," Avalanche Formation, Movement and Effects, IAHS pub. No. 162, B. Salm and H. Gubler eds.,

International Association of Hydrological Sciences, Wallingford, UK, pp. 557-570.

Buser, O., 1989, "Two years experience of operational avalanche forecasting using the nearest neighbors method," *Annals of Glaciology*, Vol. 13, pp. 31-34.

Fohn, P., W. Good, P. Bois, and C. Obled, 1977, "Evaluation and comparison of statistical and conventional methods of forecasting avalanche hazard," *Journal of Glaciology*, Vol. 19(81), pp. 375-387.

Judson, A., and B. J. Erickson, 1973, "Predicting avalanche intensity from weather data: a statistical analysis," Research Paper RM-112, U. S. Forest Service, Fort Collins, CO.

Judson, A., 1983, "On the potential use of index paths for avalanche assessment," *Journal of Glaciology*, Vol. 29(101), pp. 178-184.

LaChapelle, E., 1965, "Avalanche forecasting - A modern synthesis," International Symposium on Scientific Aspects of Snow and Ice Avalanches, IAHS Pub. No. 69, pp. 350-356.

Obled, C. and W. Good, 1980, "Recent developments of avalanche forecasting by discriminant analysis techniques: A methodological review and some applications to the Parsenn area (Davos, Switzerland)," *Journal of Glaciology*, Vol. 25(92), pp. 315-346.

Perla, R. and M. Martinelli, 1978, "Avalanche Handbook," Agriculture Handbook 489, rev. ed., USDA Forest Service, Washington, D.C..