DECISION TREES PREDICTING AVALANCHE RESPONSE: TOOLS FOR TRAINING?

Kelly Elder[1]
Rocky Mountain Experiment Station
US Forest Service
Fort Collins, Colorado

Robert E. Davis
Cold Regions Research and Engineering Laboratory
US Army Engineer Research and Development Center
Hanover, New Hampshire

ABSTRACT: This contribution shows examples of decision trees relating weather and between-storm factors to expected maximum avalanche size based on the historical records from Mammoth Mountain, California and Alta, Utah. The structure of the trees involves making binary (yes/no) decisions (e.g. 24-hour precipitation > 40 mm) to proceed from one decision node to the next until the probable response to the given conditions is reached (e.g. maximum avalanche size = 3). At each node the decision trees provide the critical factor in distinguishing one situation from another. We suggest that the decision trees in poster form may provide useful and important training guidelines for avalanche workers by increasing their historical perspective, and by emphasizing the variety and range of factors contributing to avalanche release.

KEYWORDS: Avalanche, weather factors, avalanche training

1. INTRODUCTION

In the early 1950's, Monty Atwater described and quantified a number of factors critical to avalanche release in the Wasatch Mountains of Utah (Atwater, 1954). These factors have been adopted and modified for local use around the world. Operational forecasting by an individual combines such factors with the forecaster's personal experience, knowledge and intuition. Only the elementary factors are readily transferable between forecasters. A high level of competency in forecasting depends on the experience, knowledge and intuition that may develop with many years of work in the field. Because these ingredients to a forecaster's success are difficult to transfer to others, a great deal of knowledge is lost when a seasoned forecaster leaves the field and is replaced by a novice. Tools that help new avalanche practitioners assimilate avalanche related data would help train those new to the field, however, such tools are rare.

Storm related factors and between-storm developments in the snowpack affect avalanche activity. The human forecaster takes these factors and developments into account in an experience-based assessment of the likelihood of avalanche occurrence over specific areas for a given time period. Computer software provides improved tools for handling and interpreting data to relate present observations with past conditions. Graphical representations of the quantitative analyses show structure in data sets relating weather factors and avalanche activity. Classification and regression trees produce graphical portrayals of structures found in historical observations, leading from an item by item hierarchical consideration of critical factors to the type of avalanche response expected based on existing records.

2. REGRESSION TREES AS A TRAINING TOOL

Over the past few years, we have examined extensive data sets of avalanche occurrence and coincident meteorological factors from Mammoth Mountain, California and Alta, Utah (Davis et al., 1996, 1998, 1999). We used classification and regression trees to parse the

[1]*Corresponding author address:* Kelly Elder, Rocky Mountain Experiment Station, US Forest Service, 240 W. Prospect Street, Fort Collins CO 80526-2098, voice: (970) 498-1233, fax: (970) 498-1010 e-mail: kelder@fs.fed.us

data into sets of decision rules that related the independent variables to the observed avalanche responses. The decision trees were able to explain between about 65 to 98% of the observed variance, depending on the dependent variable being described and the combination of the independent variables being used.

In examining our own results, we found that the graphical version of regression trees provided a relatively simple means to understand the complex relationships between weather and snowpack related variables used in assessing

Table 1. Weather and snow factors.

1. total snow depth
2. 24hr new snow depth
3. 48hr new snow depth
4. 72hr new snow depth
5. 24hr total precipitation
6. 48hr total precipitation
7. 72hr total precipitation
8. 24hr average wind speed
9. 48hr average wind speed
10. 72hr average wind speed
11. (24hr snowfall depth)x(24hr wind speed)^4
12. (48hr snowfall depth)x(48hr wind speed)^4
13. (72hr snowfall depth)x(72hr wind speed)^4
14. (24hr total precip.)x(24hr wind speed)^4
15. (48hr total precip.)x(48hr wind speed)^4
16. (72hr total precip.)x(72hr wind speed)^4
17. west: (24hr snowfall depth)x(24hr wind speed)^4
18. northwest: (24hr snowfall depth)x(24hr wind speed)^4
19. north: (24hr snowfall depth)x(24hr wind speed)^4
20. northeast: (24hr snowfall depth)x(24hr wind speed)^4
21. east: (24hr snowfall depth)x(24hr wind speed)^4
22. southeast: (24hr snowfall depth)x(24hr wind speed)^4
23. south: (24hr snowfall depth)x(24hr wind speed)^4
24. southwest: (24hr snowfall depth)x(24hr wind speed)^4
25. maximum air temperature
26. minimum air temperature
27. degree days
28. cumulative vapor pressure difference
29. cumulative settlement
30. starting snow depth of year
31. average snow depth increase from start of year

probability of avalanche occurrence. One of the strengths of decision trees is that results are readily interpretable, in contrast to similar relationships described by polynomial expressions, differential equations, etc. Decision trees are also ideally suited for explaining hierarchical relationships typical of processes that exhibit nonlinear behavior, such as avalanche occurrence.

3. WEATHER AND AVALANCHE DATA

Avalanche and weather observations were analyzed for ten years (1984-1993) for Alta and eight years (1989-1998; two years missing) for Mammoth, with 1895 and 890 avalanche days for each site, respectively. We used data for the maximum size of avalanche occurring on a given day as the dependent variable in this example. Independent variables included 31 separate factors that were independently measured at the field sites, were a combination of the measured factors, or were derived from the measured factors. See Table 1 for the full suite of variable used in the original studies. The examples provided in this presentation use a subset of the total number of independent variables because of space considerations.

4. RESULTS

Figure 1 shows the decision tree result for predicting/parsing the maximum size class of avalanches from the Alta data set. Figure 2 shows the Mammoth Mountain result. Discussion of these results can be found in detail in Davis et al. (1994, 1996, 1999). In this treatment we would like to consider the potential value of graphical results similar to Figure 1 and 2 as a training tool for avalanche practitioners, printed of course in a very large format. Figure 3 is a clipped branch from the Alta tree (Figure 1), which shows only the right side of the initial decision and its branches and nodes. Circles represent nodes and squares represent terminal nodes. The values inside the circles or squares represent the decision outcome, i.e. probable maximum size avalanche observed under the set of decision rules. The values under the circles and squares represent the misclassification error rate. For example, the misclassification error rate in the root node (top of the tree) is 535/1895 and the root node value is NONE. These values mean that 535 (28%) of the total 1895 cases classified in the root node are incorrectly classified as NONE, and 1360 (72%) are correctly classified. The values listed in the branches represent the decision rule for the particular split. For instance, in Figure 3 the first decision is based on the 48-hour new snow depth and the critical value is 0.16 m. This represents

Figure 1. Decision tree for maximum size class of avalanches from the Alta data.

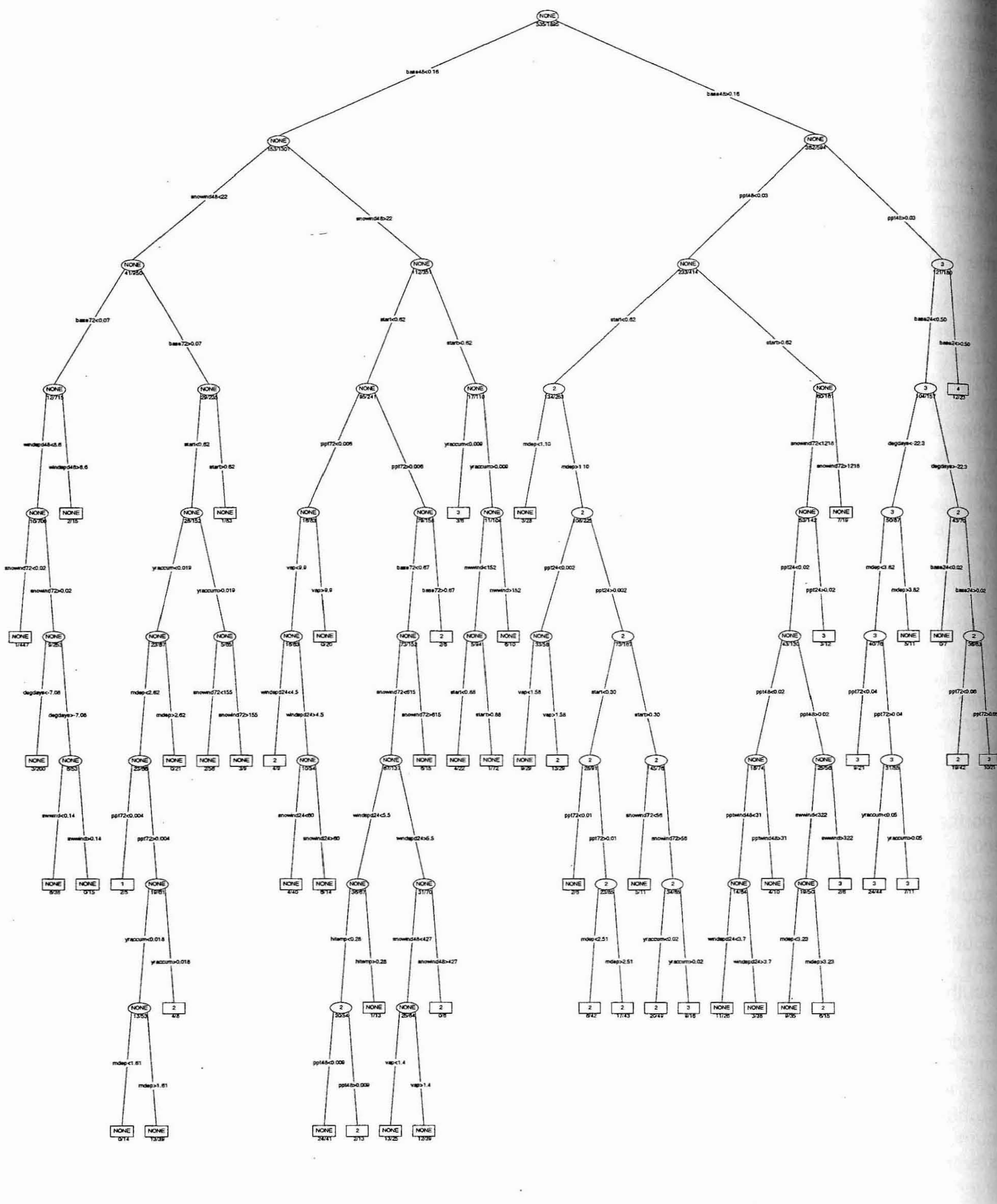


Figure 2. Decision tree for maximum size class of avalanches from the Mammoth data.

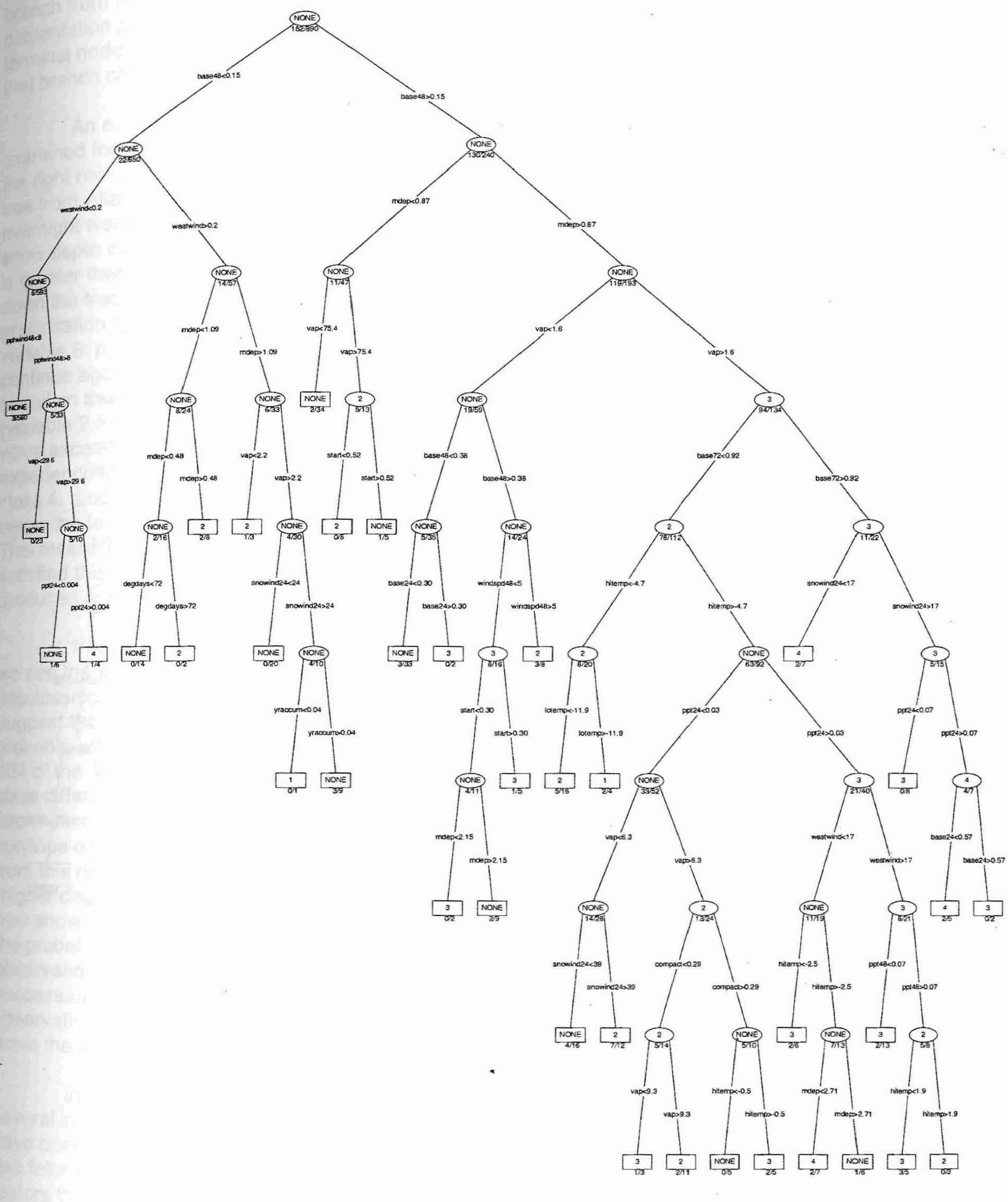
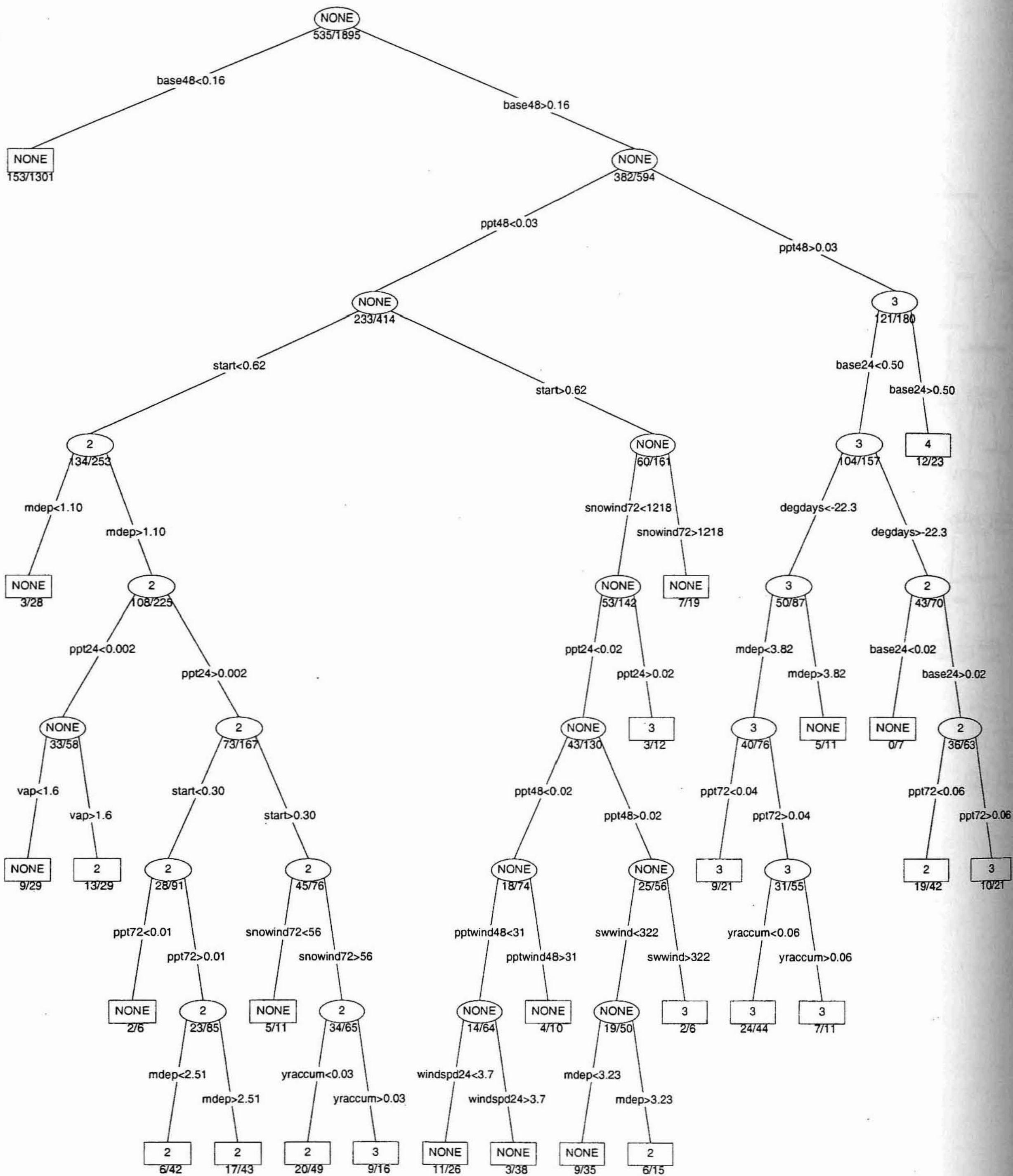


Figure 3. Right side of decision tree for Alta (Figure 1) showing most correctly modeled decisions.



variable 3 in Table 1 and is listed in Figure 3 with the variable name base48. Note that the left hand branch from the root node has been clipped for presentation purposes and the value at the terminal node is meaningless. The continuation of that branch can be seen in Figure 1.

An example from Figure 3 can be examined for illustration purposes. We will follow the right hand set of decision rules on the partial tree from Alta (Figure 1). First we examine the overnight weather record. If the cumulative new snow depth over the past 24 hours at the base site is greater than 0.16 m, then we move to the right down the tree. If the cumulative 48-hour precipitation (water equivalent of the new snow - variable 6: ppt48) is greater than 0.03 m, then we continue again to the right. The next decision is based on the previous 24-hour new snow depth (variable 2 in Table 1 and listed as base24). If the value exceeds 0.5 m, then we can expect, from the experience in the data set, avalanches up to size class 4. Looking at the misclassification error rate below the terminal node, we see the value 12/23. This means that under the previous conditions that satisfied this set of decisions, 11 of the 23 days produced avalanches of the size 4 class.

If we follow the left branch at the last split we see the size class value of 3 in the node. The misclassification error rate under that node suggest that although the size 3 events are most probable with the information given to that point, 104 of the 157 observations resulted in a size class different than 3. Clearly we need more information to reduce our uncertainty, so we continue down the tree. If we follow the right split from this node and have relatively warm days (higher degree day factor) followed by low 24-hour new snow depth, we se no avalanches (NONE) as the probable response. We find that 100% of the observations agreed with this classification, with a successful classification of seven out of seven observations. Other examples can be followed down the tree in Figure 3.

In previous the example, we hope that several important characteristics of the analysis have been noted. The decision tree structure and use follows a logical set of decisions based on factors that make physical sense to the experienced user. A less experienced user can easily follow the decision process and gain a feel for the critical factors, and their interdependence, that lead to a given result. The trees exhibit a hierarchical structure that allows one to uncover the entire sequence of decisions that would otherwise be difficult to separate. For example, in our analysis above, 24-hour new snow depth may actually be more important than 48-hour new snow depth, but the tree suggests that if the 48-hour value does not reach 0.16 m, then a different set of rules must be applied to determine a probable result.

One should also notice from examining any of the trees that many terminal nodes have the same value. Clearly we would expect this intuitively, partly because there are only six possible outcomes: four avalanche sizes (2-4 - see Davis et al., 1999) and no avalanche. However, the trees show the user the many possible scenarios that can lead to a given outcome. Again, this is potentially a valuable training lesson to the less experience avalanche worker. The trees should give them an appreciation for the complex set of processes and possible conditions that lead to avalanching, and help them make some sense of the overwhelming possibilities.

Finally, the trees are effective in detailing the set of conditions that lead to extreme events. We never have significant sample sizes to definitively predict extreme events, however, these events are of great interest to us. While we have large uncertainties and low confidence in a statistical sense, looking at the conditions that led to extreme events in the past is a valuable tool and can be added to part of a practitioner's decision process.

5. SUMMARY

We suggest that binary decision trees have potential as a training tool for avalanche practitioners because they:

- follow and demonstrate a logical thought process to the user;
- demonstrate in a simple fashion the hierarchical structure of complex relationships, and;
- detail the conditions leading to rare situations or extreme events.

These graphical models may serve as useful training tools in a formal educational setting, or as a patrol-room reminder that receives casual and occasional attention.

6. FUTURE WORK

We plan to distribute graphical models of our classification-tree results to the Alta and Mammoth Mountain ski patrols for feedback and report on this at a future workshop.

7. REFERENCES

Atwater, M.M, 1954, Snow avalanches, Scientific American, 190(1), 26-31.

Davis, R.E., and K. Elder, 1994, Application of classification and regression trees: selection of avalanche activity indices at Mammoth Mountain, Proceedings of the 1994 International Snow Science Workshop, ISSW '94, Snowbird, UT, 285-294.

Davis, R.E., K. Elder, D. Howlett and E. Bouzaglou, 1996, Analysis of weather and avalanche records from Alta, Utah and Mammoth Mountain, California using classification trees, Proceedings of the 1996 International Snow Science Workshop, ISSW '96, Banff, Alberta, 14-19.

Davis, R.E., K. Elder, D. Howlett and E. Bouzaglou, 1999, Relating storm and weather factors to dry slab activity at Alta, Utah and Mammoth Mountain, California, using classification and regression trees, Cold Regions Science and Technology, 30, 79-89.