Treed Avalanche Forecasting:
Mitigating Avalanche Danger Utilizing Bayesian Additive Regression Trees

Gail Blattenberger and Richard Fowles[1]

August, 2011

Abstract

Little Cottonwood Canyon Highway is a dead-end, two lane road leading to Utah's Alta and Snowbird ski resorts. It is the only road access to these resorts and is heavily traveled during the ski season. Fifty-seven percent of this road has been calculated to fall within known avalanche paths and the road is ranked among the most dangerous highways in the world relative to avalanche hazard. Professional avalanche forecasters monitor this road throughout the ski season in order to make road closure decisions in the face of avalanche danger. Forecasters at the Utah Department of Transportation avalanche guard station at Alta have maintained an extensive daily winter database on explanatory variables relating to avalanche prediction. Whether or not an avalanche crosses the road is modeled in this paper via Bayesian additive tree methods. Utilizing daily winter data from 1995 to 2010, results show that using Bayesian tree analysis outperforms traditional statistical methods in terms of realized misclassification costs that take into consideration asymmetric losses arising from two types of errors. Closing the road when an avalanche does not occur is an error harmful to resort owners and not closing the road when one does may result in injury or death.

1. Introduction

During the ski season, professional avalanche forecasters working for the Utah Department of Transportation (UDOT) monitor one of the most dangerous highways in the world. These forecasters continually evaluate the risk of avalanche activity and make road closure decisions. Keeping the road open when an avalanche occurs or closing the road when one does not are two errors resulting in potentially large economic losses. Road closure decisions are partly based on the forecasters' assessments of the probability that an avalanche will cross the road. In this paper, we model that probability using Bayesian additive regression

[1] Department of Economics, University of Utah, Salt Lake City, Utah. Previous research on this topic was funded, in part, by a grant from the National Science Foundation (SES-9212017).

trees (BART) as introduced in Chipman, George, and McCulloch (2010) and demonstrate that closure decisions based on BART forecasts obtain the lowest realized cost of misclassification compared with standard forecasting techniques. The BART forecasters are trained on daily data running from winter 1995 to spring 2008 and evaluated on daily test data running from winter 2008 to spring 2010. Our results generalize to decision problems that relate to complex probability models when relative misclassification costs can be accounted for.

In the following sections we explain the hazard, provide an overview of the complexity of avalanche phenomenon, describe the data, and provide an overview of the BART methodology. We then present results highlighting model selection and performance in the context of losses arising from misclassification. In our conclusion we discuss why BART methods are a natural way to model the probability of an avalanche crossing the road based on the available data and the complexity of the problem.

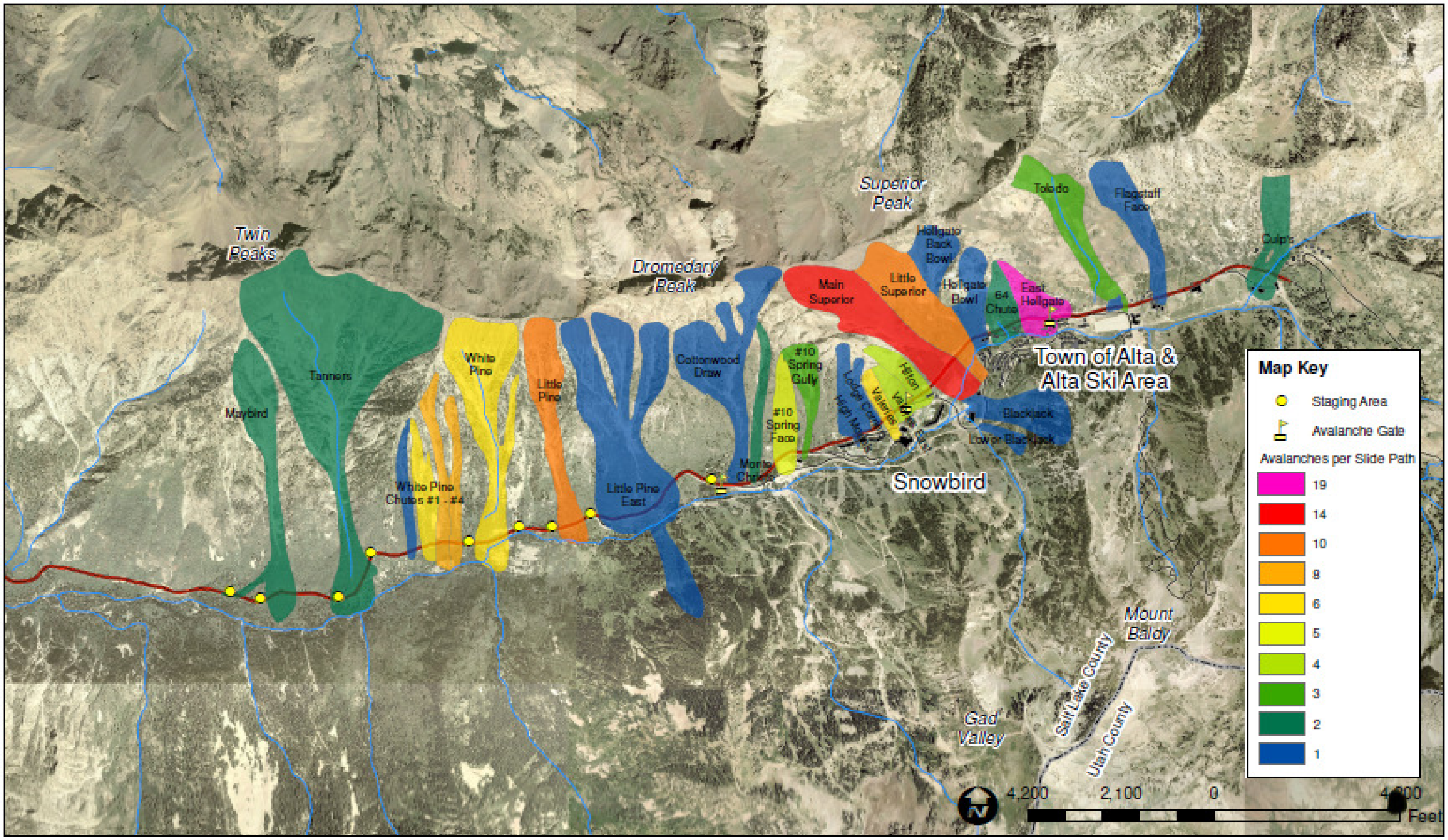
2. The Little Cottonwood Canyon Hazard

The Little Cottonwood Canyon road is a dead-end, two lane road that is the only link from Salt Lake City to two major Utah ski resorts, Alta and Snowbird. It is heavily traveled and highly exposed to avalanche danger; fifty-seven percent of the road falls within known avalanche paths. The road ranks among the most dangerous highways in the world relative to avalanche hazard. It has a calculated avalanche hazard index of 766 which compares with an index value of 126 for US Highway 550 crossing the Rockies in Colorado and an index value of 174 for Rogers Pass on the Trans Canadian Highway.[2] A level of over 100 on this index indicates that full avalanche control is necessary.

The reasons for this road's high hazard ratings are illustrated in Figures 1a, 1b, and 1c. Figure 1a shows the number of natural and controlled avalanches that affect the roadway for major avalanche paths in the canyon. There are over 20 major avalanche slide paths that cross the road. During the ski season the road is heavily utilized. Figure 1b shows daily traffic volume in the canyon for February, 2005. February is typically a month with a large number of skiers in Utah. On peak ski days, over 12,000 automobiles travel to the two resorts on the Little Cottonwood Canyon road. Figure 1c illustrates the hourly east-west traffic flow for February 26, 2005. The eastbound traffic flow is from Salt Lake City to the Alta and Snowbird ski resorts

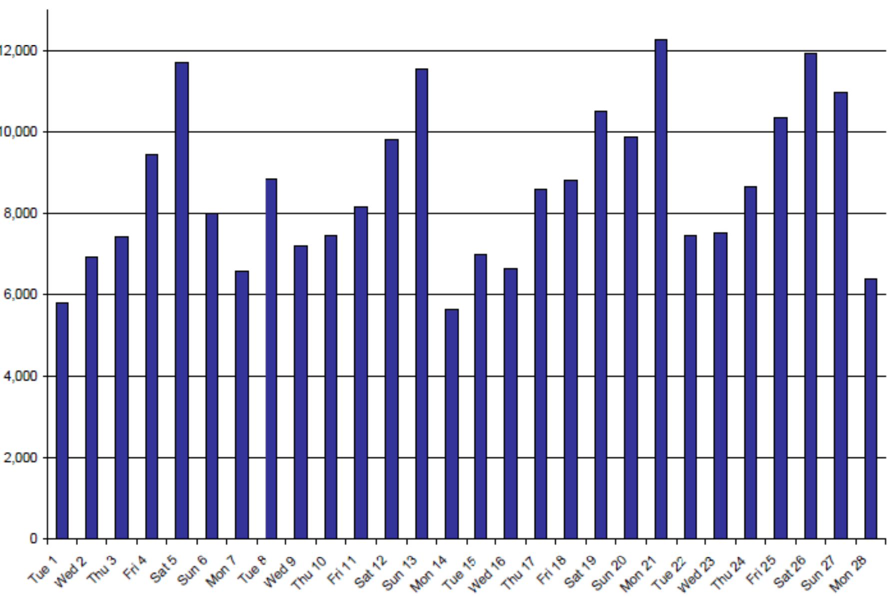
[2] See Bowles and Sandahl (1988).

Figure 1a
Natural and Controlled Avalanches by Path, 1995-2005
Little Cottonwood Canyon



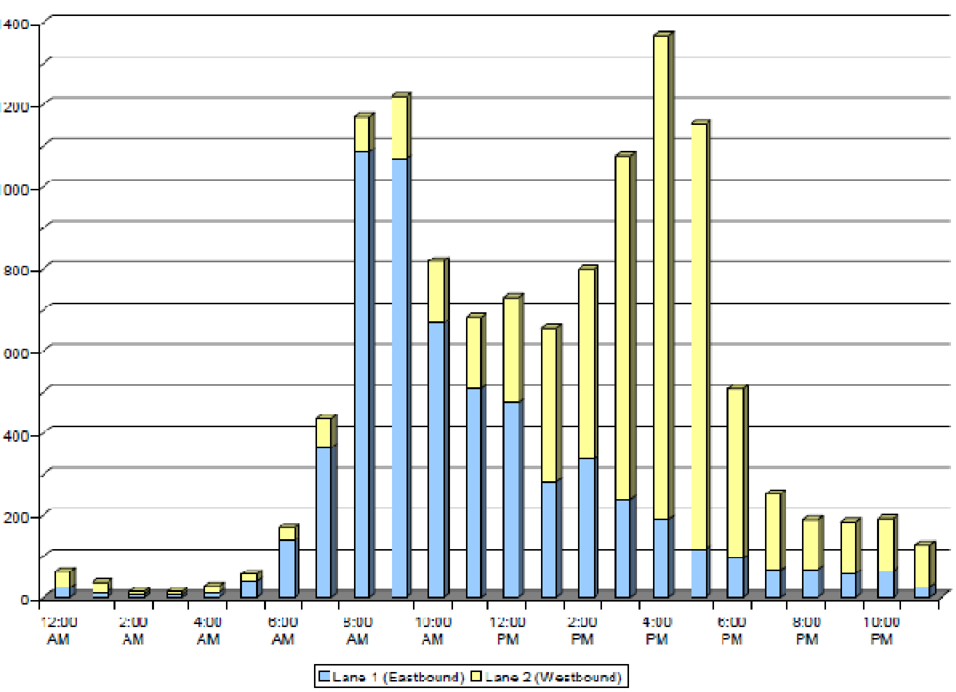
Source: Utah Department of Transportation

Figure 1b
Daily Traffic Volumes
Little Cottonwood Canyon Road, February 2005



Source: Utah Department of Transportation

Figure 1c
Hourly Traffic Volume by Direction
Saturday, February 26, 2005



Source: Utah Department of Transportation and is high in the morning hours. In the afternoon, skiers return to the city and westbound traffic flow on the road is high.

Recognition of avalanche danger along this road and attempts to predict avalanche activity began early. In 1938 the US Forest Service issued a special use permit to the Alta ski resort. One year later the Forest Service initiated full time avalanche forecasting and control.[3] By 1944, avalanche forecasters maintained daily records on weather and the snowpack. During the 1950's, forecasters began to utilize advanced snowpack instruments and meteorological information for avalanche prediction.[4] Except where noted, the measurements apply to the guard station.

Despite the fact that detailed physical measurements of climate and snowpack conditions are available, the complexity of the avalanche phenomena makes prediction difficult. Professional forecasters take into consideration multiple interactions of climate and snowpack conditions. Variables that forecasters considered in previous studies and interactions among the variables differ among forecasters, change through the season, alter across seasons, exhibit redundancy, and vary according to particular avalanche paths. For these reasons, we employ a Bayesian sum-of-trees model as presented in Chipman, George, and McCulloch (2010). Bayesian sum-of-trees models provide flexible ways to deal with high-dimensional and high-complexity problems. These problems are characteristics of avalanche forecasting and the ensemble of Bayesian trees becomes our “forecaster.’’ Sets of Bayesian forecasters contribute information that leads to a synthesized road closure decision. A closure decision is observable (the probability of an avalanche is not) and we gauge the performance of our forecasters on their subsequent realized costs of misclassification. Compared with other methods, our ensemble of Bayesian forecasters does a better job.

The next section provides a brief overview of the avalanche phenomenon and the process of avalanche forecasting.

3. Avalanche Forecasting[5]

Assuming many readers are unfamiliar with the avalanche phenomenon, we include this section to provide an introduction to avalanches and how they are forecast. Data issues relating

---

[3] See Abromeit (2004).
[4] See Perla (1991).
[5] This section is rewritten from a section in Blattenberger and Fowles (1995).

to this are held until the next section. Avalanches have climatic characteristics, and in the continental United States they are roughly characterized as maritime, transitional, and continental. Little Cottonwood Canyon is in a transitional climate zone. The mountains in Little Cottonwood Canyon run north-south and the road goes in an east-west direction. The highpoint of the road is at the eastern end in the town of Alta. Ski resorts are on the south side of the road and their slopes are primarily north-facing. Avalanche paths affecting the road are on the north side of the road and are, for the most part, south-facing slopes (Figure 1a). All avalanche paths that reach the road are monitored, but the snow study plot and most of the data collection is at the guard station at Alta.

Avalanches are complicated phenomena, and snow science and snow mechanics, which study avalanches, have developed into highly technical fields. Nonetheless many traditional, real-world forecasters have relied almost entirely on a "feel" for the situation. None rely completely on analytic models. This is partially explained because the real world conditions in which avalanche forecasts are made can differ substantially from the laboratory conditions explained in snow mechanics and snow structure science. Also the information generally available to forecasters is highly imprecise. This is partly because the information is geographically very local. There are substantial snow differences, for example, between avalanche starting zones, high on the mountains where avalanches start, and the guard station study plot where snow structure is monitored close to the road. In addition, some of the measurements themselves are imprecise. For instance two forecasters digging snow pits to appraise snow stability at the same location may come up with differing variable measurements recorded on their charts.

The data employed by forecasters is fortunately redundant, fortunate because this can compensate for imprecision. The redundancy is well illustrated by the following story.[6] Four professional forecasters at Red Mountain Pass in Colorado all had similar performances in the accuracy of their forecasts. When questioned subsequently the forecasters listed a combined total of 31 variables that they found important in their projections; individually each of the forecasters contributed less than 10 variables to the 31 total. Each focused on a collection of variables. However of the 31 variables, only one was common to all four of the forecasters.

[6] See Perla (1970).

Avalanche forecasting is not a quick decision. Hypotheses are tested and revised based on test results and on changing conditions. Characteristics of the snowpack develop over a season. Professional forecasters tend not to take breaks in the middle of a season so they will not lose contact with developments in the snowpack. The multitude of interrelated factors renders a simple forecasting model impossible. The redundancy of the information confirms our focus on the implications of the statistical model for decision ignoring estimation or parameter fit. Technical aspects of the avalanche phenomenon are explained in detail in other texts (see, for example, McClung and Schaerer (1993), or Armstrong and Williams (1986), or Perla and Martinelli (1975)). Here we simply point to some facets of the problem.

Avalanches may occur in various forms. Some are minor sluffs.[7] Although these sluffs may be deadly to an individual in the wrong location, they are not an important factor for highway closure in this particular situation. Some avalanches are deep slab avalanches transporting tons of snow down the mountain into a runout zone. It is primarily these deep slab avalanches that threaten this road. A deep slab avalanche usually, but not always, has three components. On the top there is a cohesive slab of heavy snow. On the bottom there is a bed surface along which the snow slides. The bed surface could be an ice layer resulting from a melt freeze, or even the ground. In the middle there is a weak layer between the slab and the bed surface. In addition, there is usually something that triggers any slide. Ski cuts or temperature changes can precipitate an avalanche. Explosives are used to trigger slides for control purposes, but they are not always effective. None of these features is always present and sometimes these components are difficult to identify in a particular slide.

Avalanche activity is most intensive during storms. Particular storm attributes contribute to the avalanche phenomenon. The depth of the new snow is an obvious feature. This, however, needs to be conjoined with other attributes in avalanche forecasting. The type of snow crystal affects how it will cohere to the old snow surface. The density of the snow, in terms of its water content, also affects the hazard. High density snow can cause a slab to form, especially if the density is increasing. A heavy snowfall can trigger an avalanche, particularly following a light snow. Snowfall intensity in inches per hour is another contributory factor: increasing intensity can cause instability. New snow settlement can also contribute to instability; the direction of this affect, however, can be ambiguous (Perla, 1970). High snow settlement may indicate good

[7] A minor sluff an avalanche that is small in volume of snow and loose, not cohesive.

bonding with old snow layers but it may also indicate the creation of a heavier slab. Major storms substantially increase the danger of an avalanche reaching the road.

Avalanches are not always storm events. The snowpack itself also is very important. A snowpack of sufficient depth covers terrain irregularities that would block or divert a slide. In Little Cottonwood Canyon, a 60 cm. base is thought to be a minimum depth for avalanches to occur (Perla, 1970). The snow type affects the strength of the snowpack. Snow crystals form in hundreds of identifiable types (LaChapelle, 1977) with different strengths, density, and cohesiveness with other snow layers. One finds layers of snow in pits identifiable with particular storms long after the storm occurs. Besides snowfall, surface hoar[8] forms an identifiable weak layer.

Transformations occur over time within the snowpack affecting its strength; the snowpack does not remain constant over the season. The most obvious transformation is melting: a melt-freeze, the refreezing of melted snow into an icy surface, causes a potential bed surface leading to the aforementioned deep slabs. Transformation in the snow crystals themselves also occurs as a function of the temperature gradient from the snow surface to the ground; increasing air temperature leads to a generally unstable snowpack. Thus a warm period can trigger slides.

The climate zone also affects the transformations occurring in the snowpack. Maritime climates generally have mild temperatures and high snowfall. Weak layers in the snowpack do not persist over time. On the other hand continental climate zones have low snowfall and cold temperatures. This combination results in a high temperature gradient between the ground and the snow surface. The resulting transition in the snow crystals is called depth hoar, a weak layer that can persist for a period of time. It is a notorious culprit in slab avalanches. Little Cottonwood Canyon, located in a transitional climate zone, does experience depth hoar.

All of the above conditions must be considered locally. Here by locally, we mean variations even within a slide path, not climatic zones. Temperature and snowfall vary by location and, importantly, snow is transported by wind. Terrain, wind speed, and wind direction determine the location of wind-transported snow. Slabs are often created by wind loading, the accumulation of snow deposited by wind. In addition, the aspect of a slope affects the snow transformations taking place. Major differences in perspective exist between the snow rangers at

[8] Surface hoar is frost forming on the snow surface. Surface hoar forms when the surface air is highly saturated relative to the snow surface.

the ski resorts in Little Cottonwood who principally deal with north-facing slopes, and the highway avalanche forecasters who principally monitor the south-facing slopes affecting the road.

A final factor to be considered is control activity. Control activity is roughly categorized as active or passive in nature. Passive control includes building control structures which, in Little Cottonwood Canyon, amounts to the bypass road between Alta and Snowbird on the south side of the canyon. Regulation of the structure and location of new building sites are passive control. Road closure is also termed passive control. Active control is direct action to trigger avalanches, including ski cuts, a ski path severing the snow surface and explosives. These active controls test the snow stability and when the snow is in unstable conditions release avalanches under controlled situations.

4. Data

Although the description of the avalanche phenomena is replicated from an earlier paper (Blattenberger and Fowles 1995), the data here are not. The data of the earlier study went from the 1975-76 ski season through 1992-1993. The present study uses training data running from 1995 through spring 2008 and test data from winter 2008 to spring 2010. Various sources were used for the data in the earlier study including US Department of Agriculture data tapes. The current study makes use entirely of data from the UDOT guard station. Partly as a result of recommendations made in our earlier study, additional variables were recorded and are now available from the guard station. These new variables are used here

As in the earlier study, two key variables describe closure of the road, CLOSE, and the event of an avalanche crossing the road, AVAL. Both are indicator variables and are operationally measurable constructs, a key requirement to our approach. Unfortunately, these two variables are less precise than desired. For instance the observation unit of the study is generally one day unless multiple events occur in a day, in which case CLOSE and AVAL appear in the data as multiple observations. The occurrence of an avalanche or, for that matter, a road closure is a time-specific event. It may happen, for example, that the road is closed at night for control work when no avalanches have occurred. The road is then opened in the morning and there is an avalanche closing the road. Then the road is reopened and there is another avalanche. This sequence then represents three observations in the data with corresponding data values

CLOSE = (1, 0, 0) and AVAL = (0, 1, 1). An uneventful day is one observation. If the road is closed at 11:30 at night and opened at 7:00 the following morning it is coded as closed only within the second of the two days. The variable AVAL is the dependent variable to be forecasted in this analysis. The variable CLOSE is a control variable used to evaluate model performance.

The data from the UDOT guard station is quite extensive. All of the explanatory variables are computed from the UDOT data source to reflect the factors discussed above concerning the avalanche phenomenon. However, there are no snow stratigraphy measures, or estimates of the composition of snow layers in that data. To remedy this, we made an attempt to construct a proxy for the missing stratigraphy or snow pit information. Thus the variable RELDEN measures the relative densities of the snow on past days. This would distinguish between increasing density and decreasing density indicating, although very roughly, the presence of a weak layer, which as mentioned, could lead to a slab condition. While it is a remarkable data set, the measures are sometimes ambiguous and imprecise. Fortunately, there is substantial redundancy in the data which can compensate somewhat for the imprecision.

We noted above that the variables are local, primarily taken at the Alta guard station. Measures can vary considerably even within a small location. They can vary substantially among avalanche paths and even within avalanche paths. Avalanches on these paths, illustrated in Figure 1a, affect the road.

A listing of the variables used in this study and their definitions is given in Table 1. All the variables, excepting NART, HAZARD, SZAVLAG, WSPD and NAVALLAG, were measured at the guard station. WSPD, NART, HAZARD, and SZAVLAG are new to this study. The variable HAZARD was created in response to our request in the previous paper. HAZARD is a hazard rating recorded by the forecasters. NART is the number of artificial artillery shots used. NAVALLAG is the number of avalanches affecting the road on the previous day. High values of artillery shells fired would indicate that real world forecasters believe there is instability in the snowpack requiring them to take active control measures. SZAVLAG weights these avalanches by their size rating. WSPD, wind speed, is taken at a peak location. It was not consistently available for the earlier study. The redundancy among the variables is obvious. For example, WATER = DENSITY*INTSTK, where DENSITY is the water content of new snow per unit depth and INSTK, interval stake, is the depth of the new snow. There are no snow stratigraphy measures. Monthly snow pit data were available. Snow pits are undoubtedly useful to the forecaster to learn about the snowpack, but snow pits at the Alta study plot do not reflect conditions in the starting zones of avalanche paths high up on the mountain, and monthly information was not sufficiently available. As noted above, some attempt was made to construct proxies for stratigraphy from the data available. The variable called RELDEN is the ratio of the density of the snowfall on the most recent snow day to the density of the snowfall on the second-most recent snow day. This is an attempt to reconstruct the layers in a snowpack. The days compared may represent differing lags depending on the weather. A value greater than 1 suggests layers of increasing density although a weak layer could remain present for a period of time.

We have included eighteen explanatory variables extracted from the guard station data. The large number of variables is consistent with the Red Mountain Pass story described above. The four forecasters in the story all had similar forecasting performance each using a few but differing variables. The variables included in the analysis and their definitions are given in Table 1.

Table 1 Variables used in the analysis

YEAR MONTH DAY	Date
AVAL	0-1 Avalanche crosses road
CLOSE	0-1 Road closed
TOTSTK	Total stake - total snow depth in inches
TOTSTK60	If TOTSTK greater than 60 cm.TOTSTK60 = TOTSTK - 60 in centimeters
INTSTK	Interval stake - depth of snowfall in last 24 hours
SUMINT	Weighted sum of snow fall in last 4 days weights=(1.0,0.75,0.50,0.25)
DENSITY	Density of new snow, ratio of water content of new snow to new snow depth
RELDEN	Relative density of new snow, ratio of density of new snow to density of previous storm
SWARM	Sum of maximum temperature on last three skidays, an indicator of a warm spell
SETTLE	Change in TOTSTK60 relative to depth of snowfall in the last 24 hours

WATER	Water content of new snow measured in mm.
CHTEMP	Difference in minimum temperature from previous day
TMIN	Minimum temperature in last 24 hours
TMAX	Maximum temperature in last 24 hours
WSPD	Wind speed MPH at peak location
STMSTK	Storm stake: depth of new snow in previous storm
NAVALLAG	Number of avalanches crossing the road on the previous day
SZAVLAG	The size of avalanche, this is the sum of the size ratings for all avalanches in NAVALLAG
HAZARD	Hazard rating of avalanche forecasters
NART	Number of artificial explosives used

All of the explanatory except NART, NAVALLAG, HAZARD, and SZAVLAG can be treated as continuous variables. NART, NAVALLAG, HAZARD, and SZAVLAG are integer variables; AVAL and CLOSE are factors. Descriptive statistics for these variables in the training data are given in Table 2a. The training DATA consist of 2822 observations.

Many of the variables were taken directly from the guard station data. Others were constructed. TOTSTK or total stake, INTSTK or interval stake, DENSITY or density, HAZARD or hazard rating, TMIN or minimum temperature, TMAX or maximum temperature, WSPD or wind speed, and STMSTK or storm stake came directly from the guard station weather data which is daily. TOTSTK60, SUMINT, WATER, SWARM, SETTLE, and CHTEMP were computed from the guard station weather data. NART, NAVALLAG, and SZAVLAG were constructed from the guard station avalanche data. These last three variables are not daily, but event specific and needed conversion into daily data.

SZAVLAG employs an interaction term taking the sum of the avalanches weighted by size.[9] Descriptive statistics for the 2822 observations of these variables in the training data are given in Table 2a.

[9] In computing SZAVLAG the measure which we use is the American size measure which is perhaps less appropriate than the Canadian size measure. However, a similar adjustment might be relevant.

Table 2a Descriptive Statistics for the TRAINING Data

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
AVAL	0.0	0.0	0.0	0.0361	0.0	1.0
CLOSE	0.0	0.0	0.0	0.1247	0.0	1.0
TOTSTK	0.0	33.46	63.78	61.44	90.16	159.1
TOTSTK60	0.0	25.0	102.0	104.5	169.0	344.0
INTSTK	0.0	0.0	0.0	6.076	8.00	84.0
SUMINT	0.0	0.0	7.75	15.10	24.0	122.75
DENSITY	0.0	0.0	0.0	4.694	8.333	250.0
RELDEN	0.0025	1.0	1.0	4.574	1.0	1150.0
SWARM	0.0	52.0	68.5	67.40	86.0	152.0
SETTLE	-110.0	0.0	0.0	-0.6542	0.0769	43.0
WATER	0.0	0.0	0.0	5.836	7.0	90.0
CHTEMP	-42.0	-3.0	0.0	0.0138	3.0	40.0
TMIN	-12.0	10.0	19.0	18.14	26.0	54.0
TMAX	0.0	26.0	35.0	34.58	44.0	76.0
WSPD	0.0	12.0	18.0	18.05	24.0	53.0
STMSTK	0.0	0.0	0.0	7.577 .	1.0	174
NAVALLAG	0.0	0.0	0.0	0.0698	0.0	14.0
SZAVLAG	0.0	0.0	0.0	0.203	0.0	42.0
HAZARD	0.0	0.0	1.0	0.921	2.0	4.0
NART	0.0	0.0	0.0	0.2392	0.0	23.0

The test data consist of 471 observations. Descriptive statistics for the test data are given in Table 2b.

Table 2b Descriptive Statistics for the TEST Data

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
AVAL	0.0	0.0	0.0	0.04176	0.0	1.0
CLOSE	0.0	0.0	0.0	0.1810	0.0	1.0
TOTSTK	0.0 .	24.21	74.80	61.68	90.55	141.70
TOTSTK60	0.0	1.5	130.0	106.9	170.0	300.0
INTSTK	0.0	0.0	0.0	6.385	8.000	62.0
SUMINT	0.0	0.0	8.725	15.92	23.38	87.75
DENSITY	0.0	0.0	0.	4.476	8.225	47.5
RELDEN	002105	1.0	1.0	4.073	1.0	266.7
SWARM	0.0	55.0	69.0	70.07	86.0	144.0
SETTLE	-90.0	0.0	0.0	-1.034	0.0	2.667
WATER	0.0	0.0	0.0	6.081	7.5	72.0
CHTEMP	-21.0	-4.0	0.0	0.00232	4.0	23.0
TMIN	-9.0	11.0	19.0	18.5	26.0	41.0
TMAX	0.0	27.0	35.0	35.69	44.0	72.0
WSPD	0.0	12.5	18.0	17.95	24.0	57.0
STMSTK	0.0	0.0	0.0	13.47	14.75	189.00
NAVALLAG	0.0	0.0	0.0	0.0951	0.0	8.0
SZAVLAG	0.0	0.0	0.0	0.2877	0.0	24.0
HAZARD	0.0	1.0	2.0	1.65	2.0	4.0
NART	0.0	0.0	0.0	0.4246	0.0	22.0

The data are surely not optimal. A relevant question is whether they are informative for real-world decision making. The imprecision and redundancy of the data channel our focus to the decision process itself.

5. The BART model

BayesTree is a BART procedure written by Hugh Chipman, Ed George, and Rob McCulloch. Their package, available in R, was employed here.[10] This is well documented elsewhere and here we only introduce basic concepts and the relevance to the current application.[11]

BART is an ensemble method aggregating over a number of semi-independent forecasts. Each forecast is a binary tree model partitioning the data into relatively homogeneous subsets and making forecasts on the basis of the subset in which the observation is contained. The concept of a binary tree is illustrated in Figure 2a and Figure 2b. Figure 2a presents a simple tree which explains some vocabulary. All trees start with a root node which contains all the observations in the data set. The data set is bifurcated into two child nodes by means of a splitting rule, here INTSTK $\geq$ 20. Observations with INTSTK $\geq$ 20 are put into one child node; observations with INTSTK < 20 are put into the other child node. Subsequently in this diagram, one of the child nodes is split further into two child nodes. This is based on the splitting rule, SWARM $\geq$ 50. This tree has 3 terminal nodes, illustrated with boxes here, and two internal nodes, illustrated with ellipses. The number of terminal nodes is always one more than the number of internal nodes. The splitting rules are given beneath the internal nodes. This tree has depth 2; the splitting employs two variables, INTSTK and SWARM. This partitioning of our data according to the splitting rules given here is shown in a scatter plot in Figure 2b. This scatter plot highlights the actual observations when an avalanche crosses the road. Each observation is contained in one and only one terminal node. A forecasting rule for this partition is given and the misclassification rate for each node in Figure 2b is illustrated.[12]

[10] Chipman, George, and McCulloch (2009).
[11] See Chipman, George and McCulloch (1998, 2010).
[12] This partition scores poorly but is only used to illustrate the concepts.

Figure 2a

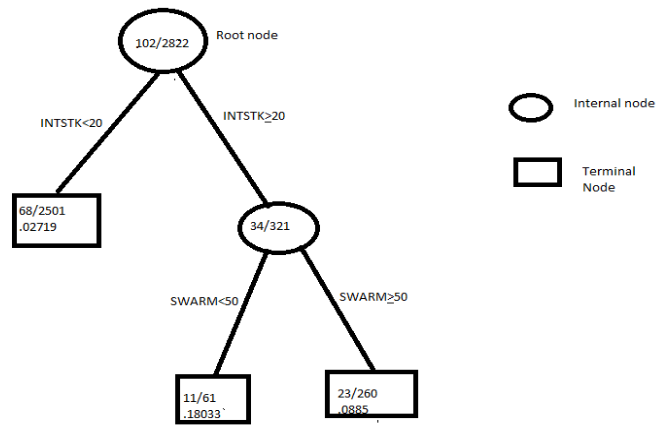
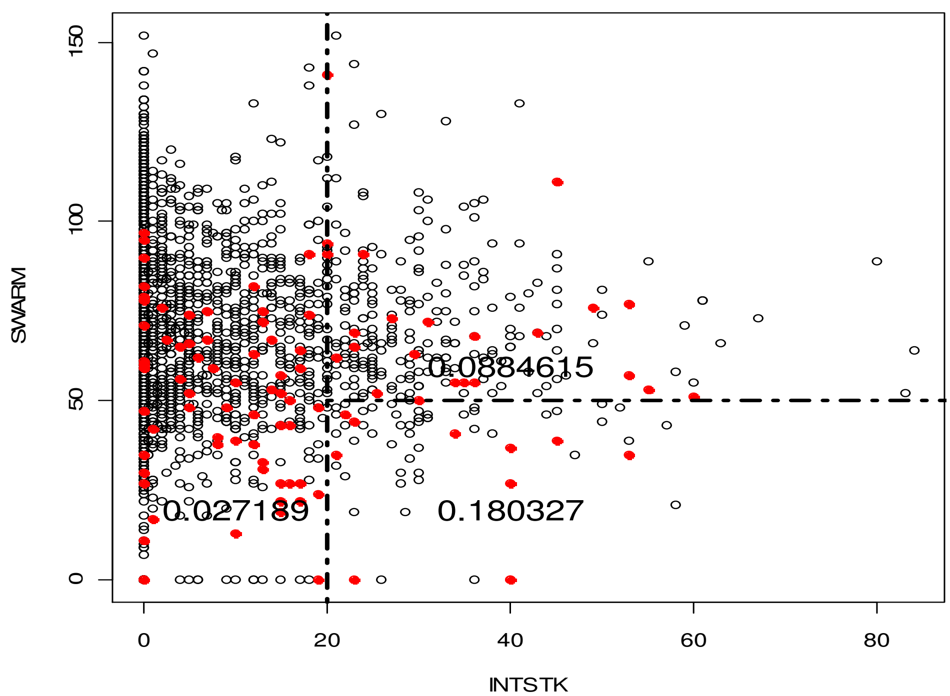


Figure 2b



The basic BART model is

$$y_i = \sum_{j=1}^{m} g\left(X_{ij} \middle| T^j, M^j\right) + u_{ij}\,;u_{ij} \sim N(0, \sigma^2)$$

where i is the observation number (i=1,…,n) and j is the j^{th} tree (j=1,…,m). Here the variable y_i is the indicator variable AVAL, indicating whether an avalanche crosses the road. Each forecaster, j, in the ensemble makes forecasts according to his own tree, T^j, and model, M^j where M^j defines the parameter values associated with the terminal nodes of T^j. It is a sum of trees model, aggregating the forecasts of the m forecasters in the ensemble, each forecaster being a weak learner.

This model seems particularly applicable to our situation. Recall the story of the four forecasters at Red Mountain Pass in Colorado. The forecasters had comparable performance. They each chose less than 10 variables out of the 36 available on which to base their forecasts. Only one of the chosen variables was common among the forecasters. Here we aggregate over an exogenous number of forecasters, each with his own tree and his own selection of variables.

The trees for the m forecasters are generated independently. Each tree is generated, however, with a boosting algorithm conditional on the other m-1 trees in a Gibbs sampling process, consequently the term semi-independent. Given the m trees generated in any iteration, the residuals are known and a new σ^2 distribution is based on these residuals. An inverse gamma distribution is used for σ^2 and the parameter distributions in the next iteration employ the σ^2 drawn from this distribution.

A Markov Chain of trees is generated for each forecaster by means of a stochastic process. Given the existing tree, T^{jk-1}, for forecaster j at iteration k-1, a proposal tree, T^*, is generated. The generation of the proposal tree is a stochastic process done according to the following steps:

- Determine the dependent variable, R^{jk}, or the "residuals" for Y conditional on the m-1 other trees, R^{jk}=Y- $\sum_{l \neq j} g(X|T^{l\gamma}, M^{l\gamma})$, where γ=k-1 if l>j, and γ=k if l<j, and determine the previous tree for this forecaster, T^{jk-1}.

- A decision is made on whether the tree will be split as defined by the probability, $a(1+d)^b$.[13]
- Given a decision to split, a decision is made on the type of split. The types of splits and their associated probabilities are: GROW (.25), PRUNE (.25), CHANGE (.4), SWAP (.1). These are described in Chipman, George, and McCulloch (1998). At the root node there is only one option, GROW. The option CHANGE is feasible only if the tree has depth greater than or equal to two. For each type of split there are a finite number of choices. GROW will occur at terminal nodes. CHANGE occurs at a pair of internal nodes, one the child of the other.
- The next decision concerns the variable on which the split is made and the splitting rule, again among a finite number of choices. The variables are equally likely. The number of potential splits depends on the variable selected, but for each variable the potential splits are equally likely.
- Given this proposal tree, a posterior distribution is determined for each terminal node based on a "regularization" prior designed to keep individual tree contributions small. Parameters are drawn from the posterior distribution for each terminal node.
- The proposal tree is accepted or rejected by a Metropolis-Hastings algorithm with the probability of accepting T^* equal to $\alpha = \min\left(\frac{q(T^{jk-1},T^*)}{q(T^*,T^{jk-1})}\frac{p(Y \mid X,T^*)p(T^*)}{p(Y \mid X,T^{jk-1})p(T^{jk-1})},1\right)$ where $q(T^{jk-1},T^*)$ is the transition probability of going from T^{jk-1} to T^* and $q(T^*,T^{jk-1})$ is the transition probability of going from T^* to T^{jk-1} . The function q() and the probabilities $P(T^*)$ and $P(T^{jk-1})$ are functions of the stochastic process generating the tree. The ratio, $\frac{p(Y \mid X,T^*)}{p(Y \mid X,T^{jk-1})}$ is a likelihood ratio reflecting the data X and Y, ensuring that the accept/reject decision is a function of the data.
- Acceptance of a tree is dependent on there being a sufficient number of observations in each terminal node of T^*.
- If the tree is accepted $T^{jk} = T^*$, otherwise $T^{jk} = T^{jk-1}$.

[13] We selected a=.95, the default value. This implies a high likelihood of a split at the root node with a decreasing probability as the depth of the tree, d, increases. The default value of b is 2. We used b=.5, however, to obtain bushier trees (trees with more terminal nodes). The story used in the text had forecasters using less than 10 variables, but at least 3.

The MCMC (Markov Chain Monte Carlo) is run for a large number of iterations to achieve convergence. The individual forecaster's trees are not identified. It is possible that trees may be replicated among forecasters in different iterations. The objective here is not parameter estimation but forecasting.

6. Results of the BART Application

6.1 Splitting Rules

Before discussion of the performance of our forecasting model we look at some of our choices concerning the BART process. First we specify the number of trees or, as we called them above, the number of forecasters. We used for comparison purposes 50, 100, and 200 trees, preferring 50 trees. We also had to specify the parameters of the splitting rule for the tree generating process, P(split) = $a(1+d)^b$. We selected a=.95, the default value. This implies a high likelihood of a split at the root node with a decreasing probability as the depth of the tree, d, increases. The parameter b relates to the bushiness of the tree. First we examined the average number of terminal nodes per iteration, and per tree for the first 3000 iterations after the break-in period. These are given in Table 3.

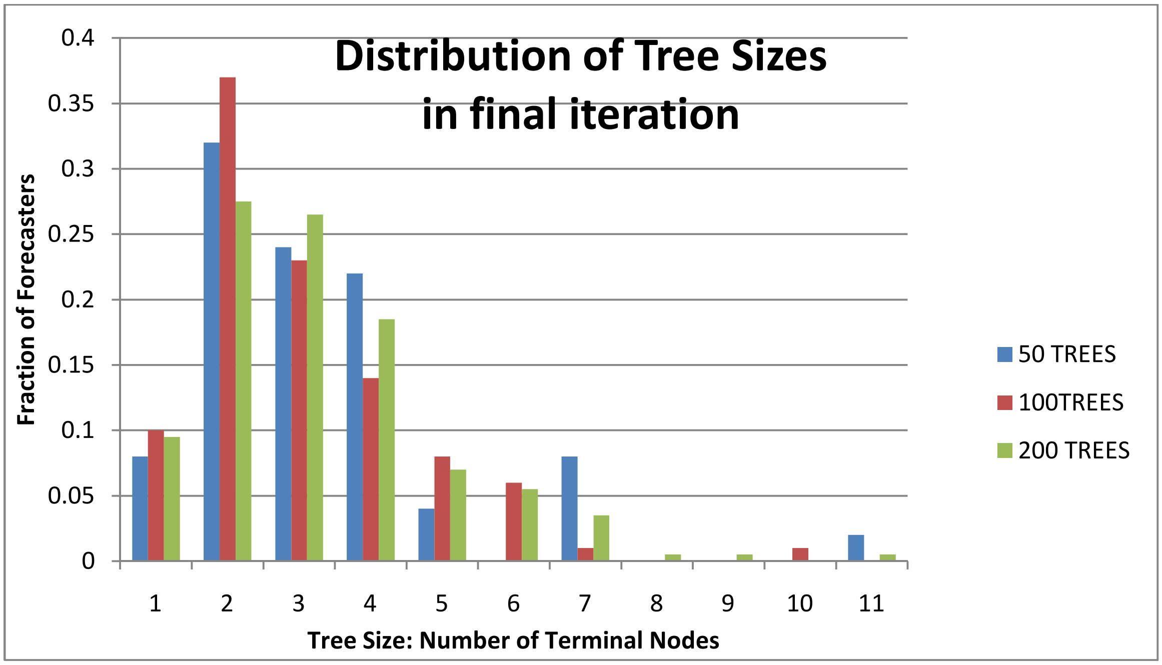
Table 3
Average Tree Size per Tree and Iteration
Given ntree = number of trees and split parameter b

	Power=b								
ntree	0.5	0.6	0.7	0.8	0.9	1	1.2	1.5	2
50	3.429	3.168	3.112	2.974	2.914	2.804	2.741	2.500	2.338
100	3.222	3.153	3.048	2.961	2.901	2.837	2.659	2.499	2.315
200	3.222	3.166	3.077	2.980	2.872	2.802	2.659	2.500	2.317

The choice of 50 trees and b=0.5 yields an average of 3.4 terminal nodes. We aim to be consistent with the Perla story on Red Mountain Pass. While Table 3 describes the average number of trees, there is substantial variation among the forecasters in any single iteration. The frequency distribution of tree sizes among forecasters within the last iteration is pictured in

Figure 3. While tree size may vary substantially for any specific forecaster across iterations, the last iteration should be representative for post break-in iterations.

Figure 3



This is consistent with each forecaster in the story making his decision based on less than 10 variables.

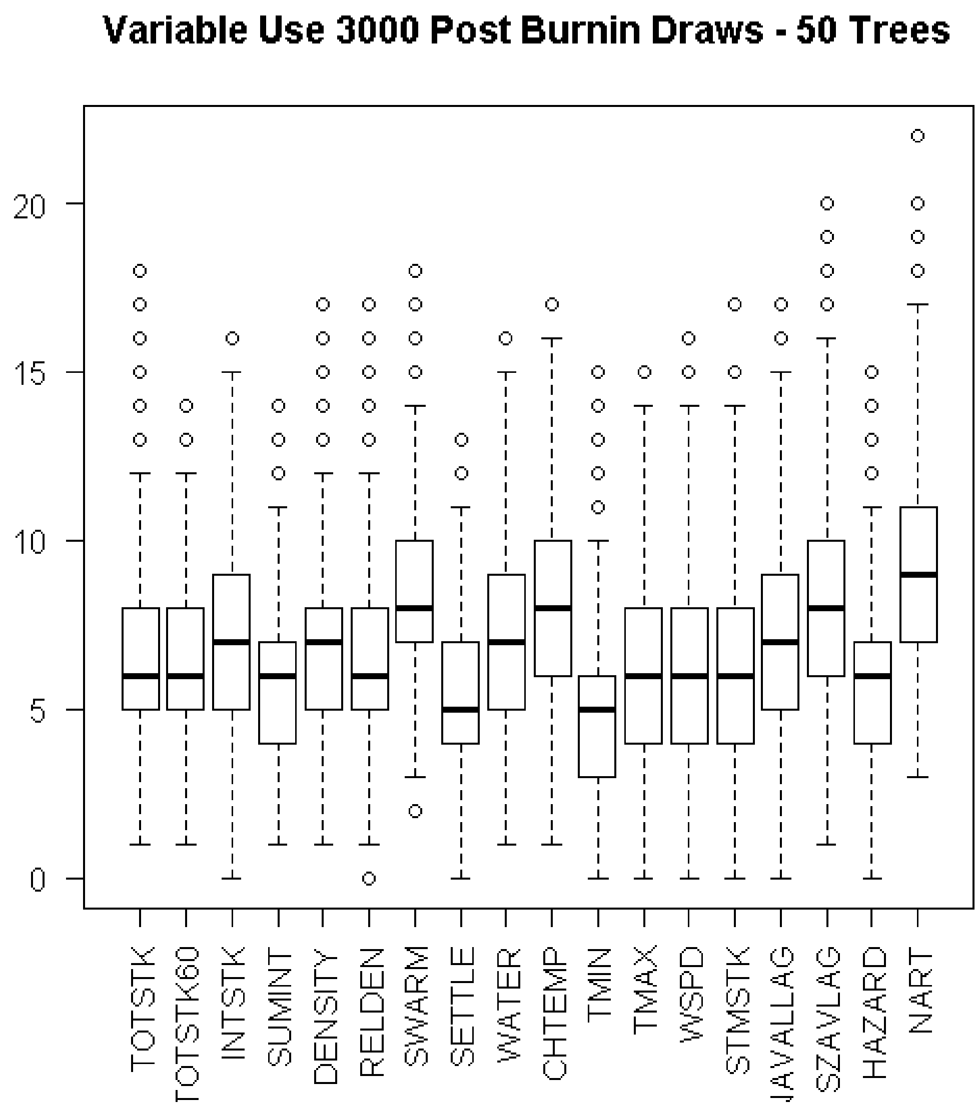
6.2 Variable choice

We noted in the Red Mountain Pass example that the four forecasters had only one variable in common in spite of the fact that their forecasts were comparably accurate. An interesting comparison is the variable choice among our forecasters. This is illustrated in Figure 4 for 50 forecasters showing a box-and-whisker plot for variable use among 50 forecasters in 3000 post burn-in iterations. The vertical axis gives the number of forecasters using each variable. A value of 50 would indicate a variable used by every forecaster. No such variable exists. All variables on average were used by at least five forecasters. This conforms again with our comments on the redundancy of the variables and the Red Mountain Pass story.

The most commonly used variable was NART, the number of artificial explosives used and the least commonly used variable was HAZARD, a hazard rating of the forecasters. It may be that the decision to use artificial explosives more accurately reflects the forecasters' evaluation of avalanche hazard than the hazard rating itself.

SWARM, the presence of a warm period, and CHTEMP, the change in temperature, are also prominent variables, as is SZAVLAG, the recent occurrence of many large avalanches. There are numerous indicators of snow depth and storm size for forecasters to choose amongst. There is redundancy between TOTST and TOTSTK60 relating to the depth of the snow pack. Similarly redundancy exists among INTSTK, SUMINT, and STMSTK, measures of storm activety, as well as among DENSITY, WATER, WSPD, and SETTLE and among temperature variables TMIN, TMAX, CHTEMP, and SWARM. All are selected by some forecasters with similar frequencies but no one dominates. Although Figure 4 illustrates variable choice for 50 forecasters, similar results were obtained for 100 and 200 forecasters.

Figure 4



6.3 Realized cost of misclassification

We now turn to the forecast performance of the BART model. A common measure of forecast performance is root mean squared error, RMSE. The RMSE values for the avalanche forecasting models are as follows:

Table 4 Root Mean Square Error for Test Period

Linear	Logit	BART 50	BART 100	BART 200	Guard Station
0.165	0.165	0.161	0.163	0.162	0.397

The BART model with 50 forecasters wins on this criterion. We noted earlier, however, that all forecasting errors are not equivalent. This issue needs to be addressed in evaluating the forecasts.

We assume that the forecasters act to minimize expected losses associated with their road closure decision. The asymmetric loss function is:

$$\text{Loss} = k^*p + q$$

In this loss function, p represents the fraction of the time that an avalanche crosses the road and it is open and q represents the fraction of the time that an avalanche does not cross the road and it is closed. The term k is a scale factor that represents the relative cost of failing to close the road when an avalanche occurs to the cost of closing the road when an avalanche did not occur. Both p and q are observable, while k is not. The decision rule to minimize expected loss implies an implicit cutoff probability, $k^* = 1/(1+k)$, such that the road should be closed for probabilities greater than k^* and kept open for lower probabilities. In previous work we found a value of $k=8$ to be consistent with the historical performance of the avalanche forecasters and in line with revenue losses to the resorts relative to loss of life estimates.[14]

To evaluate BART model performance we examine the realized cost of misclassification (RCM) or loss. It is calculated as a function of the cutoff probability. Figure 5a compares RCMs for linear, logit, and BART predictions (from a 50 tree model). We also plot the experts' performance over the testing period as a horizontal line at 0.22. BART performance is nearly uniformly lower than other models for cutoff probabilities from 0.1 to 0.6. Figure 5b adds BART models with 100 and 200 forecasters for comparison purposes.

[14] Details are in Blattenberger and Fowles (1994, 1995). UDOT data indicates that, on average, there are 2.6 persons per vehicle, 2.5 of which are skiers. Of these skiers, 40% are residents who spend an average of $19 per day at the ski resorts (1991 dollars). 60% tended to be non-residents who spent an average of $152 per day (1991). We estimate a road closure results in a revenue loss in 2005 of over $2.25 million per day based on average traffic volume of 5,710 cars during the ski season.

All of the BART models outperform the logit and the linear models. They also outperform the guard station decisions although the guard station decisions are immediate and are subject to certain legal constraints.[15]

[15] The road must be closed while artificial explosives are used.

Figure 5a

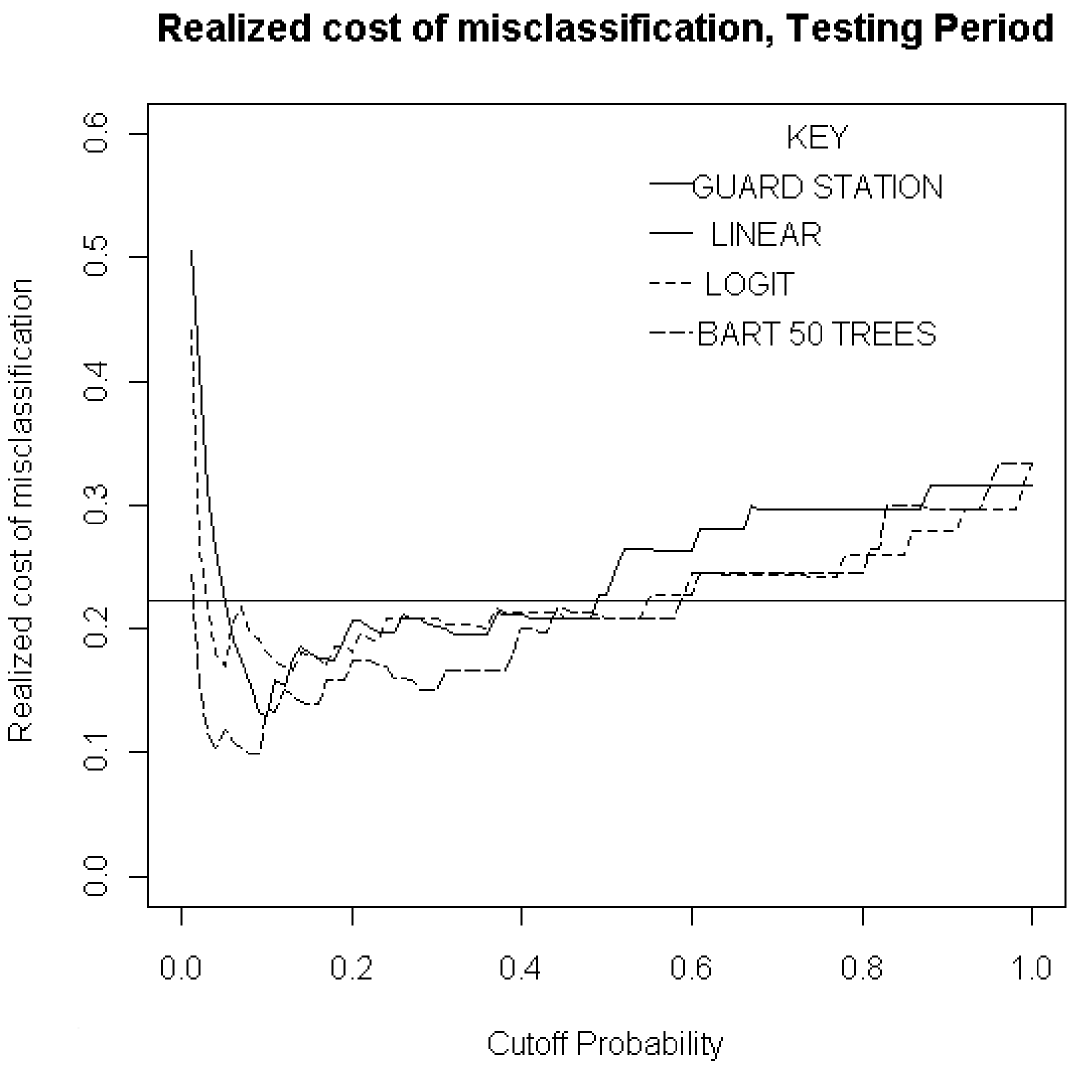
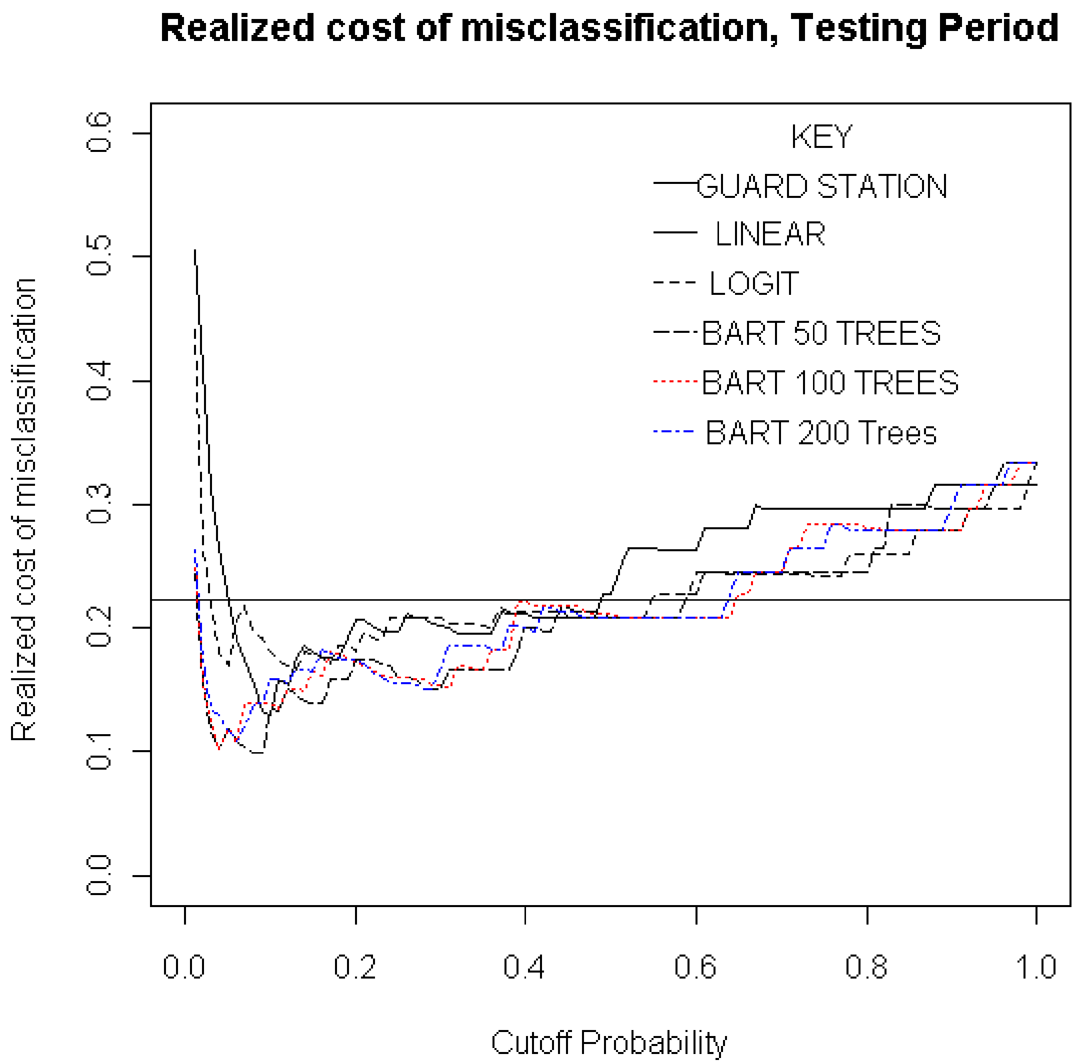

Realized cost of misclassification, Testing Period
KEY
GUARD STATION
LINEAR
LOGIT
BART 50 TREES
Realized cost of misclassification
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.0
0.2
0.4
0.6
0.8
1.0
Cutoff Probability

Figure 5b



6. Conclusion

This paper illustrates the advantage of using Bayesian additive regression trees in a real-world decision making context. By summing over many models, each contributing a small amount of information to the prediction problem, BART achieves high out-of-sample performance as measured by a realistic cost of misclassification. The philosophy behind BART is to deal with a complicated issue – analogous to sculpting a complex figure – by "adding and subtracting small dabs of clay."[16] This method seems well suited to the problem of avalanche prediction where individual professional forecasters develop an intuitive approach and cannot rely on a single analytic model.

[16] From Chipman, George, Lemp, and McCulloch (2010).

References

Abromeit, D. (2004) “United States Military for Avalanche Control Program: A Short History in Time.” Proceedings of the International Symposium on Snow Monitoring and Avalanches.

Armstrong, E., and Williams, K. (1986) *The Avalanche Book*, Golden, CO: Fulcrum.

Blattenberger, G., and Fowles, R. (1994) “Road Closure: Combining Data and Expert Opinion,” In *Case Studies in Bayesian Statistics*, New York: Springer-Verlag.

Blattenberger, G., and Fowles, R. (1995) “Road closure to mitigate avalanche danger: a case study for Little Cottonwood Canyon”, *International Journal of Forecasting*, 11, 159-174.

Bowles, D., & Sandahl, B. (1988) “Avalanche Hazard Index for Highway 210 - Little Cottonwood Caynon Mile 5.4 to Mile 13.1.” mimeographed.

Chipman, H, George, E.I, and McCulloch, R.E. (1998) “Bayesian CART Model Search”, *Journal of the American Statistical Association,* Vol. 93,No.443, 935-948

Chipman, H., George, E.I., and McCulloch, R.E. (2009) BART: Bayesian Additive Regression Trees, http://CRAN.R-project.org/package=BayesTree.

Chipman, H., George, E.I., and McCulloch, R.E. (2010) “BART: Bayesian Additive Regression Trees,” *Annals of Applied Statistics*, 4, 266-298.

Chipman, H., George, E.I, Lemp, J, and McCulloch, R.E. (2010) “Bayesian flexible modeling of trip durations,” *Transportation Research Part B*, 44, 686-698.

LaChapelle, E.R. (1977) *A Field Guide to Snow Crystals*, Seattle, WA: University of Washington Press.

LaChapelle, E.R. (1980), “Fundamental Processes in Conventional Avalanche Forecasting,” *Journal of Glaciology,* 26, 75-84.

McClung, D., and Schaere, P. (1993) *The Avalanche Handbook*, Seattle, WA: The Mountaineers.

Perla, R. (1970) “On contributory in Avalanche Hazard Evaluation,” *Canadian Geotechnical Journal,* 414-419.

Perla, R., and Martinelli, Jr., M. (1975) *The Avalanche Handbook*, Washington, DC: US Forest Service.

Perla, R. (1991) “Five Problems in Avalanche Research,” In *CSSA Symposium*.