Chapter 11
Avalanche Forecasting: Using Bayesian Additive Regression Trees (BART)

Gail Blattenberger and Richard Fowles

11.1 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. This paper models that probability using Bayesian additive regression trees (BART) as introduced in Chipman et al. (2010a, b) and demonstrates that closure decisions based on BART forecasts obtain the lowest realized cost of misclassification (RCM) 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. The results generalize to decision problems that relate to complex probability models when relative misclassification costs can be accounted for.

The following sections explain the problem, the data and provide an overview of the BART methodology. Then, results highlighting model selection and performance in the context of losses arising from misclassification are presented.

Previous research on this topic was funded, in part, by a grant from the National Science Foundation (SES-9212017).

G. Blattenberger (✉)
University of Utah, 11, 981 Windsor St, Salt Lake, UT 84105, USA
e-mail: gail.blattenberger@economics.utah.edu

R. Fowles
University of Utah, 11, 260 S. Central Campus Drive; Orson Spencer Hall, RM 343, Salt Lake, UT 84112-9150, USA
e-mail: fowles@economics.utah.edu

J. Alleman et al. (eds.), *Demand for Communications Services - Insights and Perspectives*, The Economics of Information, Communication, and Entertainment,
DOI: 10.1007/978-1-4614-7993-2_11,

The conclusion discusses 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.

11.2 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 travelled and highly exposed to avalanche danger; 57 % 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.[1] A level of over 100 on this index indicates that full avalanche control is necessary.

There are over 20 major avalanche slide paths that cross the road. During the ski season, the road is heavily utilized. Figure 11.1a 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 or return to the city. Figure 11.1b 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 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.[2] By 1944, avalanche forecasters maintained daily records on weather and the snowpack. During the 1950s, forecasters began to utilize advanced snowpack instruments and meteorological information for avalanche prediction.[3] 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,

[1] See Bowles and Sandahl (1988).

[2] See Abromeit (2004).

[3] See Perla (1991).

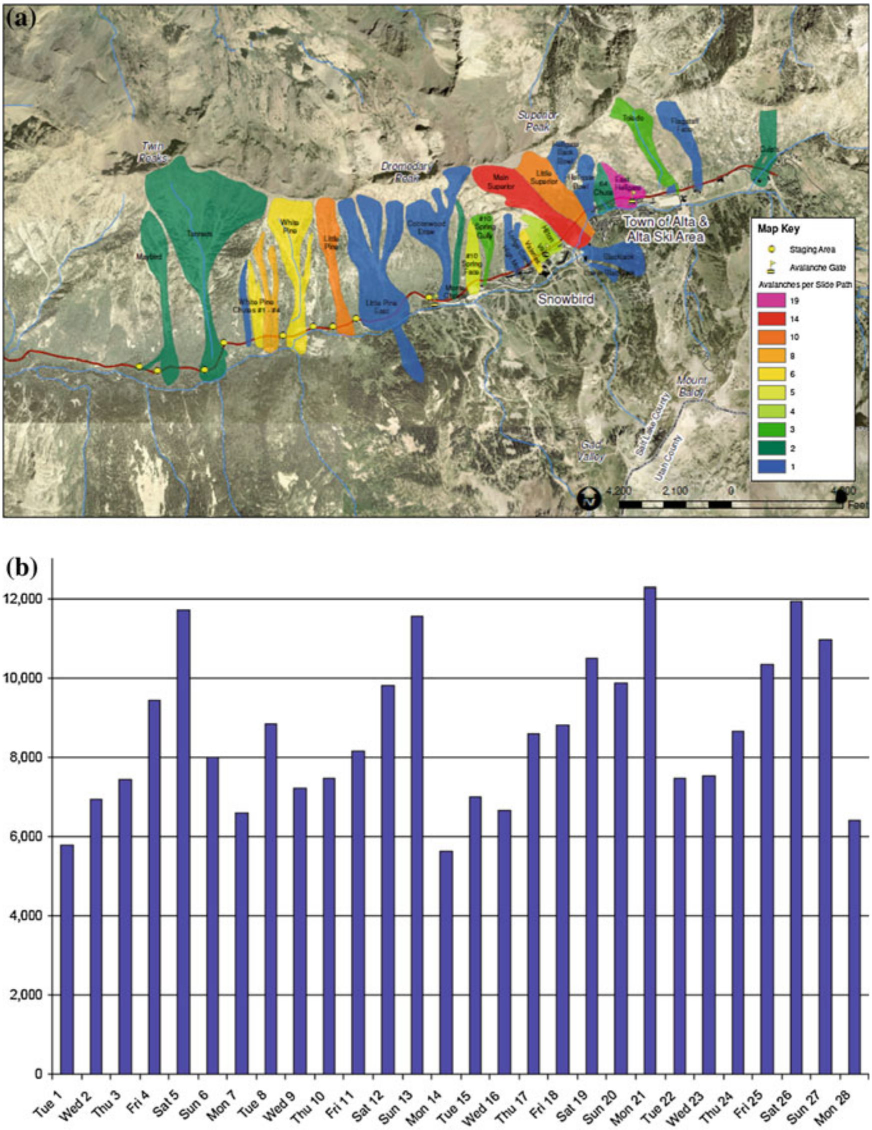


Fig. 11.1 **a** Natural and controlled avalanches by path, 1995-2005, little cottonwood canyon. **b** Daily traffic volumes little cottonwood canyon road, February 2005. **c** Hourly traffic volume by direction saturday, February 26, 2005

and vary according to particular avalanche paths. For these reasons, a Bayesian sum-of-trees model as presented by Chipman et al. (2010a, b) is employed. Bayesian sum-of-trees models provide flexible ways to deal with high-dimensional and high-complexity problems. These problems are characteristics of avalanche

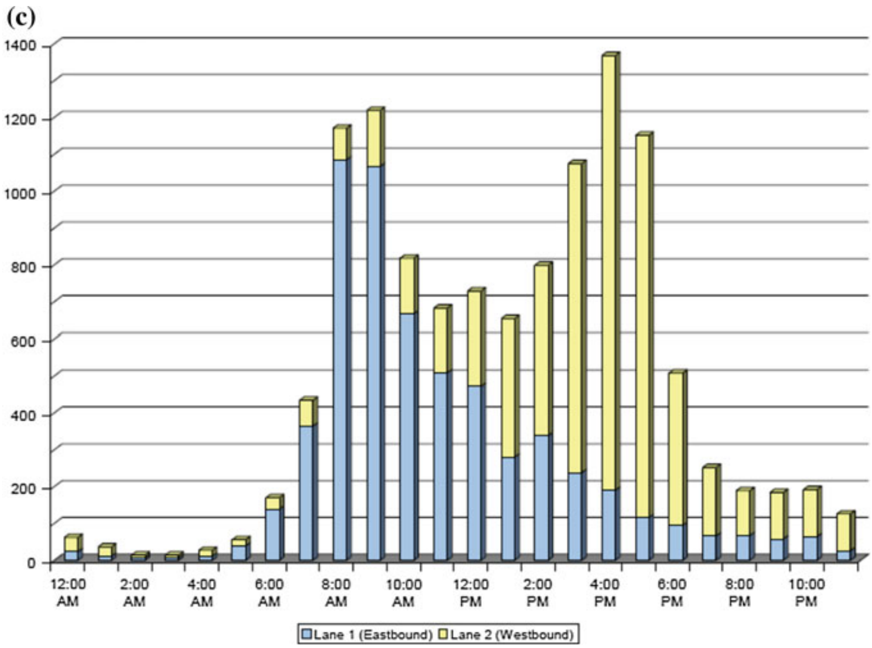


Fig. 11.1 continued

forecasting and the ensemble of Bayesian trees becomes the "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 RCM. Compared with other methods, the ensemble of Bayesian forecasters does a better job.

11.3 Data

An earlier study was performed on the road closure decision in Little Cottonwood Canyon (See Blattenberger and Fowles 1995, 1994). The data of the earlier study, however, went from the 1975–1976 ski season through 1992–1993. The present study uses training data running from 1995 to 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 the 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 2 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 are quite extensive. All of the explanatory variables are computed from the UDOT data source to reflect the factors concerning the avalanche phenomenon. 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.

A listing of the variables used in this study and their definitions is given in Table 11.1. 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 the request in the previous paper (Blattenberger and Fowles 1995). 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. SZAVLAG weights these avalanches by their size rating. High values of number of artillery shells fired, NART, would indicate that real-world forecasters believe that there is instability in the snowpack requiring them to take active control measures. 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. Only 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

Table 11.1 Variables used in the analysis

VARIABLE NAME	VARIABLE DEFINITION
YEAR MONTH DAY	Forecast Date (year, month, day)
AVAL	Avalanche crosses road : 0=no, 1=yes
CLOSE	Road closed: 0=open 1=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 ofprevious storm
SWARM	Sum of maximum temperature on last three skidays, an indicator of a warmspell
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 inNAVALLAG
HAZARD	Hazard rating of avalanche forecasters
NART	Number of artificial explosives used

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.

The data employed by forecasters are fortunately redundant,[4] fortunate because this can compensate for imprecision. The redundancy is well illustrated by the following story. 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. Of the 31 variables, however, only one was common to all four of the forecasters (Perla 1970).

Eighteen explanatory variables extracted from the guard station data were included. The large number of variables is consistent with the Red Mountain Pass

[4] The word redundant more generally than correlation is used in this paper. This indicates when several variables are designed to measure the same thing or may be functions of each other.

story described. The four forecasters in the story all had similar forecasting performance each using a few but differing variables.

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 11.2. The training DATA consist of 2,822 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.[5] Descriptive statistics for the 2,822 observations of these variables in the training data are given in Table 11.2.

The test data consist of 471 observations. Descriptive statistics for the test data are given in Table 11.3.

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.

11.4 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.[6] This is well documented elsewhere and only basic concepts and the relevance to the current application are introduced here.[7]

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 Fig. 11.2a and b. Figure 11.2a 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

[5] 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.

[6] Chipman et al. (2009).

[7] See Chipman et al. (1998, 2010a, b).

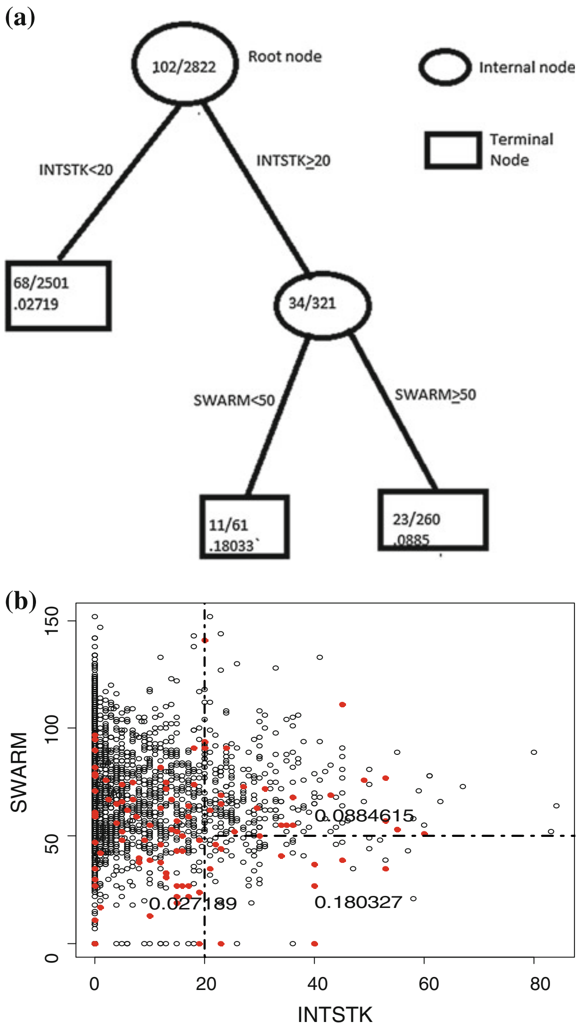
Table 11.2 Descriptive statistics used for the TRAINING data

Variables	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	1,150.0
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

Table 11.3 Descriptive statistics used for the TEST data

Variables	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	0.02105	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

Fig. 11.2 Presents a simple tree which explains some vocabulary



data set is bifurcated into two child nodes by means of a splitting rule, here INTSTK ≥20. Observations with INTSTK ≥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 ≥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 the data according to the splitting rules given here is shown in a scatter plot in Fig. 11.2b. 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 Fig. 11.2b is illustrated.[8]

The basic BART model is

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

where i is the observation number ($i = 1, \ldots ,n$) and j is the jth tree ($j = 1, \ldots ,m$). 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 this 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 31 available on which to base their forecasts. Only one of the chosen variables was common among the forecasters. Here, aggregate is 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, $R^{jk} = Y - \sum_{l \neq j} g\left(X|T^{ly}, M^{ly}\right)$, where $1 = k-1$ if $l > j$, and $1 = \mathrm{k}$ if $l < j$.

A decision is made on whether the tree will be split as defined by the probability, $a(1 + d)^b$.[9]

---

[8] This partition scores poorly but is only used to illustrate the concepts.

[9] The default value for a, 0.95, is selected. 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. However, $b = 0.5$ is used 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.

Given a decision to split, a decision is made on the type of split. The types of splits and their associated probabilities are: GROW (0.25), PRUNE (0.25), CHANGE (0.4), SWAP (0.1). These are described in Chipman et al. (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 $\alpha = \min\left(\frac{q(T^{jk-1},T*)}{q(T*,T^{jk-1})}\frac{p(Y|X,T*)p(T*)}{p(Y|X,T^{jk-1})p(T^{jk-1})}, 1\right)$ 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, $\frac{p(Y|X,T*)}{p(Y|X,T^{jk-1})}$ 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 $T^{jk} = T^{jk-1}$.

The Markov Chain Monte Carlo (MCMC) 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.

11.5 Results of the BART Application

11.5.1 Break-in Period

The MCMC is known to converge to a limiting distribution under appropriate conditions, and a number of iterations are discarded to insure the process has settled down to this distribution. It is not established, however, when this convergence is reached. The MCMC history of forecasts, for a number of dates in the training data in Fig. 11.3a–d, is examined. In these figures, a break-in period of 5,000 iterations was used, with 50 trees or forecasters. Each point, in the history, is the aggregation of the 50 forecasters for that iteration. A number of days were selected to see how the process does in differing conditions. Although there is

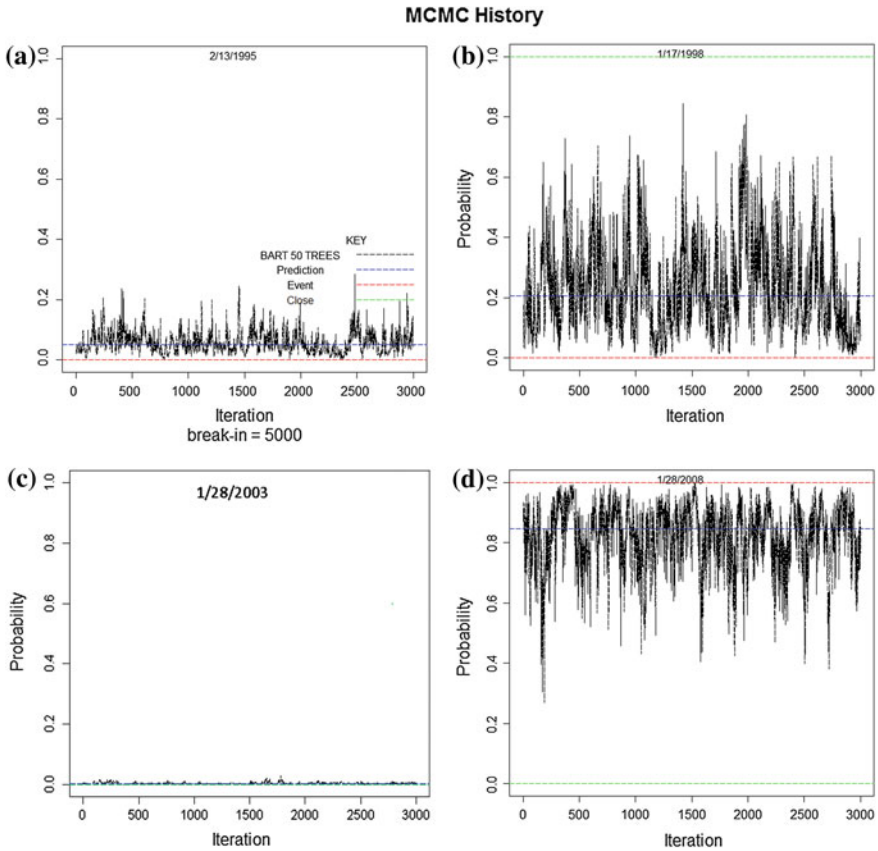


Fig. 11.3 The MCMC history of forecasts for a number of dates in the training data

variation among the iterations, the forecasts "mix" well in that there is no functional trend among the iterative forecasts. The MCMC standard error varies among the dates selected but is relatively uniform within each date.

11.5.2 Splitting Rules

Before discussion of the performance of the forecasting model, some of the choices concerning the BART process were looked at. First the number of trees or as called above the number of forecasters is specified. For comparison purposes, 50, 100, and 200 trees were used. Also, the parameters of the splitting rule for the tree-generating process, $P(\text{split}) = a(1 + d)^b$, had to be specified. The default value a = 0.95 was selected. This implies a high likelihood of a split at the root node with a decreasing probability as the depth of the tree, d, increases.

Table 11.4 Average tree size per tree and iteration. Given n tree = number of trees and split parameter b Power = b

n tree	0.5	0.6	0.7	0.8	0.9	1	1.2	1.5	2
50	3.421	3.162	3.118	2.974	2.872	2.848	2.697	2.500	2.338
100	3.283	3.146	3.061	2.961	2.904	2.822	2.673	2.499	2.315
200	3.252	3.196	3.056	2.967	2.881	2.797	2.657	2.500	2.317

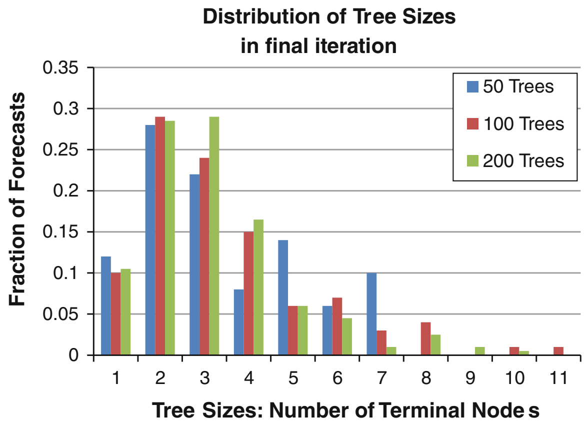


Fig. 11.4 The frequency distribution of tree sizes among forecasters within the last iteration is pictured

The parameter b relates to the bushiness of the tree. First, the average number of terminal nodes per iteration and per tree for the first 3,000 iterations after the break-in period was examined. These are given in Table 11.4.

The choice of 50 trees and $b = 0.5$ yields an average of 3.4 terminal nodes. To be consistent with the Perl, a story on Red Mountain Pass was examined. While Table 11.4 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 Fig. 11.4. While tree size may vary substantially for any specific forecaster across iterations, the last iteration should be representative for post break-in iterations.

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

11.5.3 Variable Choice

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 the forecasters. This is illustrated in Fig. 11.5 for 50 forecasters showing a box-and-whisker plot for variable use among 50 forecasters in 3,000 postburn-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

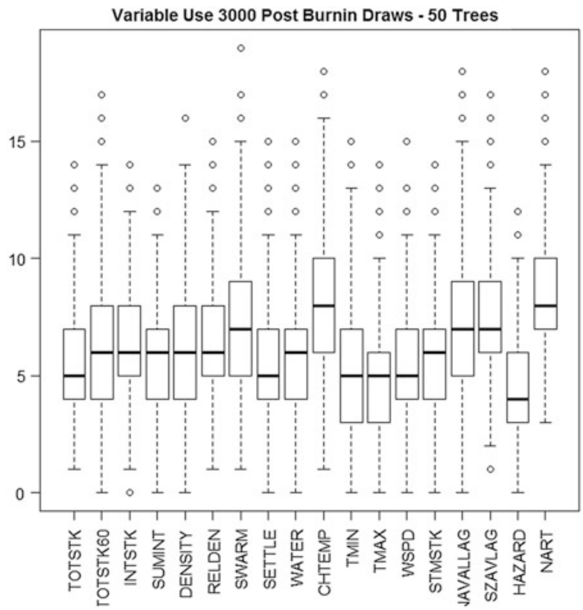


Fig. 11.5 Illustrates variable choice for 50 forecasters

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 noted 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 among. 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 activity, as well as among DENSITY, WATER, and RELDEN, among WSPD, and SETTLE and among the temperature variables TMIN, TMAX, CHTEMP, and SWARM. All are selected by some forecasters with similar frequencies but no one dominates. Although Fig. 11.5 illustrates variable choice for 50 forecasters, similar results were obtained for 100 and 200 forecasters.

11.6 Realized Cost of Misclassification

Before turning to forecast performance in the test period, Fig. 11.3a–d illustrate some relevant concepts. These figures illustrate the history of postbreak-in iterations on particular dates, a jagged black line. The actual event that occurred is

Table 11.5 Root mean square error for test period

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

shown by a dotted line at zero or one; the road closure is given by a dashed line again at zero or one. The forecast for each date is the average of the iterative values shown by a dotted dash line.

On 13 February 1995, the model predicted a low probability of an avalanche crossing the road; this was correct, but the road was closed. On 1 January 1998, the model predicted a moderate probability of an avalanche crossing the road; the road was closed, but again there was no avalanche. On 27 December 2003, the model predicted a low probability of an avalanche crossing the road; the road was not closed, and there was no avalanche. On 28 January 2008, the model predicted a high probability of an avalanche crossing the road; the road was not closed, but there was an avalanche.

We now turn to the forecast performance of the BART model in the test period, a common measure of forecast performance is root mean squared error, RMSE. The RMSE values for the avalanche forecasting models are as follows (Table 11.5).

The BART model with 100 forecasters wins on this criterion. However, as noted earlier, all forecasting errors are not equivalent. This issue needs to be addressed in evaluating the forecasts.

If one assumes 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 Blattenberger and Fowles (1994, 1995) 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.[10]

[10] Details are in Blattenberger and Fowles (1994, 1995). UDOT data indicate 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). Sixty percent tended to be nonresidents who spent an average of $152 per day (1991). 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.

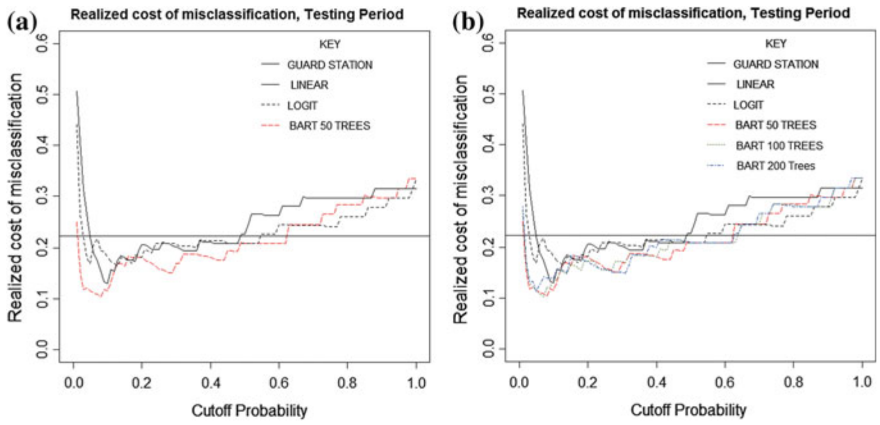


Fig. 11.6 Compares RCMs for linear, logit, and BART predictions (from a 50 tree model)

To evaluate BART model performance, the RCM or loss was examined. It is calculated as a function of the cutoff probability. Figure 11.6a compares RCMs for linear, logit, and BART predictions (from a 50 tree model). The experts' performance over the testing period as a horizontal line at 0.22 was also plotted. BART performance is nearly uniformly lower than other models for cutoff probabilities from 0.1 to 0.6. Figure 11.6b adds BART models with 100 and 200 forecasters for comparison purposes.

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.[11]

11.7 Conclusion

This paper illustrates the advantage of using BART 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" (Chipman et al. 2010a, b). 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.

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

References

Abromeit D (2004) United States military for avalanche control program: a short history in time. In: Proceedings of the international symposium on snow monitoring and avalanches

Blattenberger G, Fowles R (1994) Road closure: combining data and expert opinion. In: Gatsonis et al. (eds.) Case studies in bayesian statistics, Springer, New York

Blattenberger G, Fowles R (1995) Road closure to mitigate avalanche danger: a case study for little Cottonwood canyon. Int J Forecast 11:159–174

Bowles D, Sandahl B (1988) Avalanche hazard index for highway 210—little Cottonwood Canyon mile 5.4–13.1. mimeographed

Chipman H, George EI, McCulloch RE (1998) Bayesian CART model search. J Am Stat Assoc 93(443):935–948

Chipman H, George EI, McCulloch RE (2009) BART: bayesian additive regression trees, http://CRAN.R-project.org/package=BayesTree

Chipman H, George EI, McCulloch RE (2010a) BART: bayesian additive regression trees. Ann Appl Stat 4:266–298

Chipman H, George EI, Lemp J, McCulloch RE (2010b) Bayesian flexible modeling of trip durations. Transp Res Part B 44:686–698

LaChapelle ER (1980) Fundamental processes in conventional avalanche forecasting. J Glaciol 26:75–84

Perla R (1970) On contributory in avalanche hazard evaluation. Can Geotech J 7:414–419

Perla R (1991) Five problems in avalanche research. In: CSSA symposium